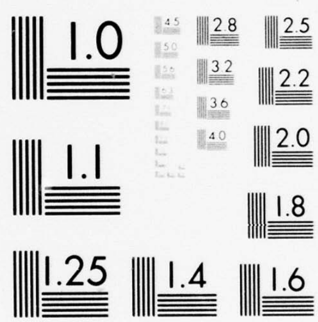
AD-A057 972 PENNSYLVANIA STATE UNIV UNIVERSITY PARK APPLIED RESE--ETC F/G 6/10
DISCRIMINATION OF NOISE-LIKE SOUNDS INVOLVING MULTIPLE INTERACT--ETC(U)
MAY 78 D W MARTIN N00017-73-C-1418

UNCLASSIFIED TM-78-171 NL

1 OF 2

AD
A057972


1.0
4.5
5.0
5.6
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

LEVEL II (12) l

(6) DISCRIMINATION OF NOISE-LIKE SOUNDS INVOLVING MULTIPLE INTERACTING FEATURES.

(9) Master's thesis,

(10) Douglas W./Martin

(12) 162p.

Technical Memorandum
File No. TM 78-171
(11) 25 May ~~25,~~ 1978
Contract No. N00017-73-C-1418
(15)
(14) TM-78-171

Copy No. 1

The Pennsylvania State University
Institute for Science and Engineering
APPLIED RESEARCH LABORATORY
Post Office Box 30
State College, PA 16801

APPROVED FOR PUBLIC RELEASE
DISTRIBUTION UNLIMITED

NAVY DEPARTMENT

NAVAL SEA SYSTEMS COMMAND

391 007

mt

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TM 78-171	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) DISCRIMINATION OF NOISE-LIKE SOUNDS INVOLVING MULTIPLE INTERACTING FEATURES		5. TYPE OF REPORT & PERIOD COVERED M.S. Thesis, August 1978
		6. PERFORMING ORG. REPORT NUMBER TM 78-171
7. AUTHOR(s) Douglas W. Martin		8. CONTRACT OR GRANT NUMBER(s) N00017-73-C-1418
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Pennsylvania State University Applied Research Laboratory P. O. Box 30, State College, PA 16801		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Sea Systems Command Department of the Navy Washington, D. C. 20362		12. REPORT DATE May 25, 1978
		13. NUMBER OF PAGES 161 pages & figures
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified, Unlimited
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited, per NSSC (Naval Sea Systems Command), 6/28/78

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

auditory perception
signal detection
discrimination
sound

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The objective of experiments reported here was to analyze the effect of feature interactions on discrimination performance with complex, noise-like sounds. Sixteen pairs of laboratory-generated sounds were used. Acoustic features composing the stimuli included octave bands of noise and amplitude modulation of noise bands by a 10-Hz square wave. Within each sound pair, signals differed by one or more "dichotomous" features, features present in one signal and not in the other. → next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

20. ABSTRACT (Continued)

The thesis presents a feature extraction model for discrimination, incorporating concepts from signal detection theory and auditory information processing. Interactions between features are hypothesized wherein the presence of one feature affects the detectability of another.

Experiments using the "modified threshold procedure" were conducted with five subjects to test the hypothesis. Subjects were presented with two signals, one of which appeared as the probe in white noise. Measured quantities were the signal-to-noise ratio to respond and the probability of a correct discrimination decision.

Experiments dealt with such topics as the relative importance of multiple dichotomous features in discrimination, the effect of an irrelevant feature adjacent in frequency to a dichotomous noise band, the role of amplitude modulation as an irrelevant feature, and the role of the background noise as a confusion parameter. Some hypothesized interactions were observed in the data while others were not.

ACCESSION for

NTIS

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

ABSTRACT

The objective of the experiments reported in this thesis was to analyze the effect of feature interactions on discrimination performance with complex noise-like sounds. Noise-like sounds were defined as any sounds, other than speech or music, which could potentially convey information to a listener. Discrimination between such sounds plays an important role in many industrial settings.

Sixteen pairs of laboratory-generated sounds were used for the experiments. Acoustic features composing the stimuli included octave bands of noise as well as amplitude modulation of noise bands by a 10-Hz square wave. Within a given sound pair, signals differed by one or more "dichotomous" features, features present in one signal and not in the other. Discrimination performance was studied under various combinations of "fixed" or irrelevant features as well as several conditions involving multiple dichotomous features.

The thesis presents a feature extraction model for discrimination performance which incorporates concepts from the theory of signal detectability and information theory as applied to auditory processing. The model hypothesizes a number of interactions between features in which the presence of one feature affects the detectability of another.

Experiments using the "modified threshold procedure" were conducted to test the predictive capabilities of the model, with the experiments involving five graduate student subjects. They were presented with two signals, one of which appeared as the probe in a white noise

background. The signal-to-noise ratio was increased slowly until a subject was willing to commit to a terminal discrimination decision. The important measured quantities were the mean signal-to-noise ratio to respond and the probability of a correct response.

These data were used to compute the detectabilities of dichotomous noise bands and the Weber fractions associated with dichotomous modulation. Values of these quantities were then compared among experiments in order to determine the extent to which feature interactions affected discrimination performance. Important conclusions included the following:

1. A single fixed noise band does not affect the discrimination of a dichotomous feature when the dichotomous feature is either a broadband noise or amplitude modulation. This applies whether the fixed band is adjacent or far removed in frequency from the dichotomous feature. However, discrimination performance was degraded in several experiments involving more than one fixed noise band.

2. Amplitude modulation as an irrelevant feature was found to degrade performance under three different conditions of dichotomous features.

3. When signals involved several dichotomous features, one of which was amplitude modulation, the discrimination decision was dominated by this feature.

4. When signals involved two dichotomous noise bands, the perceived difference between them was a unified percept rather than

two separate bands. However, the extent to which each band contributed to the perceived difference cannot be determined from the data.

5. In addition to acting as a masker, the background noise in many cases acted as a confusion parameter to the extent that it sounded like the features to be detected.

Results were compared with those predicted by the model, showing that some hypothesized interactions occurred, while others did not. Several anomolies in the data were discussed, and areas for further study were suggested.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

The author gratefully acknowledges the contributions of Dr. C. P. Janota in almost every phase of this work. Most of the equipment used for the experiments was designed by Dr. Janota. In addition, his technical assistance and suggestions regarding the choice of stimuli and analysis of results were invaluable to the completion of this thesis. This individual was also responsible for the design of the computer programs used to process the thesis text.

Also, the author expresses his appreciation for the suggestions and encouragement of Dr. R. W. Farwell, his thesis advisor. Furthermore, the assistance of fellow graduate students P. T. Cornell, in the review of experimental procedures, and A. H. Barbour, in the preparation of figures, is greatly appreciated and warrants recognition.

Finally, the assistance rendered by the Applied Research Laboratory of The Pennsylvania State University under contract with the Naval Sea Systems Command is greatly appreciated. The work was sponsored by the Naval Air Systems Command (AIR 370).

CHAPTER I

INTRODUCTION

1.1 General

Since the advent of the industrial revolution, the presence of noise has become increasingly important in man's environment. Noise is in some cases a pollutant, and in others, a provider of useful information. As a pollutant, it may result in stress or hearing loss after prolonged exposure. It may also serve to prevent the acquisition of useful information by masking speech or warning signals in industrial or other environments.

However, noise is at times useful in that it can convey valuable information. Noise from machinery can inform the operator if his machine is working correctly, and which of several operations are being performed (Zagoruyko and Voloshina, 1969). For example, a motorist whose car uses manual transmission relies on the sound of the engine to know when he should shift gears. A lathe operator uses the sound of his machine to judge the cutting rate. This thesis deals with the information conveyed by noise-like sounds, and in this context, noise will be defined as any sound which can potentially provide a listener with information in a form other than speech or music. If the term "noise" is used in any other context, its meaning will be made clear.

The problem of how a person discriminates between noise-like sounds is a difficult one which has received little attention in the literature. If two complex sounds are very dissimilar, the discrimination is easy, but if they are quite similar, what information does a listener use to discriminate between them? Furthermore, if two complex sounds differ in a number of ways, i.e., along a number of dimensions or features, which of these features is most important to the listener in deciding which sound he is hearing?

1.2 Statement of the Problem

This thesis presents experiments whose results will help to explain how a human discriminates between complex nonspeech sounds. The results of psychoacoustic experiments will be presented, and these will be analyzed in terms of a model employing concepts from information theory and the theory of signal detectability. It will be assumed that the discrimination involves a process of feature extraction within the auditory system, and that the extracted features are matched with stored representations of auditory patterns. The experiments presented here involve the discrimination of sounds which differ along one, two, and three or more dimensions.

Janota (1977) has shown that the detection of a dichotomous feature is in many cases both a necessary and sufficient condition for the discrimination of sounds differing in the presence or absence of that feature. When two sounds differ along a number of dimensions, it is hypothesized here that some of these dimensions carry relevant

information, some redundant information, and some features may serve to mask or otherwise cause confusion with relevant features. In such cases, it is probably not necessary for the observer to detect all of the dichotomous features in a sound pattern in order to make a discrimination.

This thesis is therefore an attempt to determine how various features, specifically bands of noise and amplitude modulation of noise bands, interact in complex sound discrimination. It will deal with such questions as: What are the detectabilities of the various features? How are these detectabilities affected by the presence of other features? And how do the detectabilities of the various features combine to produce an overall detectability of the difference between sounds? A simple linear combination of feature detectabilities is probably not appropriate for the task at hand since one must be concerned with the relevance of information conveyed by the various dichotomous dimensions. That is, the detection of any single feature may provide enough information to make a discrimination. On the other hand, the interaction of features may provide cues or create other features, such as auditory beats when dealing with tonal stimuli, which would in some cases be more detectable than any of the features singly. Also, some features will occur only when others are present; i.e., certain dimensions may be correlated with one another. These may provide the listener with clues that there is enough information available to decide if the dichotomous features are present or absent.

However, as mentioned earlier, other components of the signal may also act to either mask the dichotomous feature, or they may provide information which is confusing to the listener. The background noise presented on all experimental trials may also act as a confusion parameter above and beyond the role of masker by sounding like the features the subject wishes to detect. It must be emphasized at this point that the internal representation of features within the auditory system may not be perfectly correlated with the acoustic features of the signals. However, it is hoped that analysis along these lines will lead to predictions of which features are most relevant in making discrimination decisions. If some hierarchy of importance and detectability of features can be established, a more descriptive model of the discrimination process should result.

1.3 Approach

The experiments performed here involve sixteen different sound pairs in which subjects were asked to identify which of two sounds was presented in a white noise background. These were laboratory-generated sounds composed of various fixed and dichotomous features. Acoustic components of the sounds included octave bands of noise centered at several frequencies as well as periodic amplitude modulation of noise bands. The experimental paradigm used has been termed the "Modified Threshold Technique" (Janota, 1977), and it involves a sequential classification task in which signal-to-noise ratio increases with time on a given trial. A subject responds when

confident of a terminal decision. That is, he has the option of either responding or waiting for more information if the relevant features are masked by the noise. By knowing the signal-to-noise ratio and percent correct responses for sounds with various feature combinations, the detectabilities of the various components can be determined. By knowing the level at which subjects respond, it may be determined which features are most relevant to the given discrimination problem. Experimental results will be compared with those predicted by a generalized pattern recognition model incorporating concepts from information theory and the theory of signal detectability.

Before presenting experiments designed to elicit information about how an observer discriminates between complex sounds with multiple interacting features, a review of relevant literature on complex sounds is given in Chapter II. This should help to clarify some reasons for the particular choice of sounds to be tested and the choice of experimental paradigms used. Very little relevant literature is available dealing with the discrimination of noise-like sounds. A great deal of work has been done in the areas of speech discrimination (Mattingly et al., 1971; Stevens and House, 1972) and the discrimination of complex tonal stimuli (Green, 1958; Nordmark, 1972). Some of this literature is also reviewed in Chapter II, but as many investigators point out, the discrimination of speech and nonspeech stimuli may involve quite different processes (Mattingly et al., 1971; Webster et al., 1973). One must therefore be very careful in generalizing results

obtained with speech to areas of nonspeech sounds.

Chapter III presents some theoretical considerations which lead to hypotheses concerning interacting features and their relation to the discrimination task. A review of literature concerning the theory of signal detectability and information theory is presented in the context of how these theories pertain to feature detection and information content of features. Chapter III concludes with a discussion of a feature extraction and pattern recognition model (Reed, 1973).

Chapter IV begins with a list of hypotheses which follow naturally from the theoretical discussions and recognition model in the previous chapter. The chapter then describes the experiments conducted to test these hypotheses. The noise-like sounds under study, which involve bands of noise and amplitude modulation of noise bands, are presented. Methods and procedures used for the training of subjects are discussed as is the experimental paradigm called the "modified threshold procedure." The chapter concludes with a brief discussion of how the data are to be analyzed. The important parameters are the signal-to-noise ratio, SNR, necessary to obtain a response and the probability of a correct response, P(C).

Chapter V presents the results obtained with university student subjects using the modified threshold procedure. Results are compared with the model of feature extraction discussed in Chapter III, and with the hypotheses concerning multiple feature interactions.

Experimental biases as well as factors affecting subjects' decision criteria which must be taken into account when using the modified threshold procedure are also presented in this chapter. If such biases and factors affecting criteria are properly analyzed, results which are both reliable and valid are obtainable using the procedures outlined. Thus, such results should lead to a better understanding of the relevant features in the discrimination of complex stimuli, and of the detectabilities of such features under various conditions.

CHAPTER II

REVIEW OF INVESTIGATIONS ON COMPLEX SOUND DISCRIMINATION

2.1 General

Despite the importance of understanding how humans discriminate among various complex noise-like sounds, comparatively little work has been done in this area. There have been a few studies concerned with the detection and/or identification of engine sounds (Webster et al., 1969; Fidell et al., 1974). A few investigators have looked at complex sound discrimination as it applies to sonar operator training (Corcoran et al., 1968; Corcoran et al., 1970; Woodhead et al., 1973), and a few studies have dealt with detection and discrimination of marine sounds (Stallard and Leslie, 1974; Janota, 1977). In addition, two investigations have analyzed perceptual confusions among multi-dimensional stimuli (Webster et al., 1973; Howard and Silverman, 1975). However, most work with complex sounds and their ability to convey information to observers have dealt with speech (Stevens and House, 1972; Mattingly et al., 1971; Tartter and Eimas, 1975) and with complex tonal sounds (Green, 1958; Licklider and Green, 1961; Plomp, 1967; Nordmark, 1972). Most investigators believe that discrimination in the speech mode is a fundamentally different process than that in nonspeech modes, and therefore,

results obtained with speech cannot be generalized to include other complex sounds (Webster et al., 1973; Mattingly et al., 1971).

Results of studies on speech discrimination will be briefly discussed in this chapter since basic concepts of auditory feature extraction originated with such studies. Some findings on complex tonal stimuli along with the small quantity of literature dealing with discrimination among noise sources will be reviewed in the present chapter as a foundation for the experiments to be presented in later chapters.

2.2 Studies Involving Complex Tonal Signals

Discrimination studies involving complex, nonspeech sounds most frequently involve stimuli which sound like tones. These include signals made up of sinusoidal components, square waves, or signals having a fundamental frequency and various added formants. In fact, so numerous are these studies in comparison with those involving noise signals that the term "complex sound" has often been used incorrectly to describe only complex tones. The primary interest of this thesis is the discrimination of noise-like sounds. However, much of the literature dealing with tonal stimuli is relevant to this study and will be reviewed below.

It has been hypothesized that the auditory system acts in some ways like a set of tunable filters whose bandwidths and center frequencies are adjusted by the listener to match the task at hand (Swets et al., 1962; Creelman, 1960). One relevant question which

must be raised in the context of complex sound discrimination is: given a signal with components widely separated in frequency, can the listener tune his filter bank to monitor several components simultaneously? That is, can he use information contained in many components, or is the hearing mechanism limited to processing only the most detectable feature of such a signal? Green (1958) investigated this problem using multi-component signals made up of pure tones separated in frequency by various amounts and masked by white noise. Green analyzed his data in the context of three different mathematical models, all of which were possible extentions of the critical band concept first proposed by Fletcher (1940), and discussed by Scharf (1972). The data gave strongest support for a statistical summation model, suggesting that several critical bands may be linearly combined by the auditory system. The detectability of each of the multi-component signals was greater than that of any of the separate components, implying that the listener is capable of monitoring several bands simultaneously. Green stated that his results implied that "the auditory mechanism may change the appropriate parameters of the analysis process to match the signal to be detected" (Green, 1958).

Further support for a statistical summation model for predicting the detectabilities of multi-component tones is obtained in a study by Licklider and Green (1961). In their study, observers were asked to detect a signal composed of sixteen sinusoids. As in Green's 1958

experiments, the signals were partially masked by white noise, and the subjects' task was to indicate if the signal was present in the interval. Thus, these were not discrimination tasks but merely detection tasks. In the investigation (Licklider and Green, 1961), the detectabilities of each component as well as that of the combined signal were determined. Results were compared with two models. The statistical summation model states that,

$$d'_{overall} = \left(\sum_{i=1}^{n} d'^{2}_{i}\right)^{1/2} \quad (1)$$

where d'_i is the detectability of the ith component of the signal.

This implies that the components are uncorrelated, and that they add as orthogonal signal vectors. The other model asserts that the overall detectability of the signal will be no greater than that of the most detectable component. The obtained results agreed within 0.5 dB with those predicted by the summation model, whereas the discrepancy between the data and the "most detectable feature" model was greater than 6 dB.

The results obtained in the two studies reported above apply only to cases in which the signal components are widely separated in frequency exceeding a critical bandwidth. If two components are closely spaced in frequency, the degree of correlation between them depends on the shape which one assumes for the auditory filter (Green, 1958). The relations between effective critical bandwidth and various assumptions about its shape in the auditory mechanism have been discussed in a paper by Swets et al. (1962).

It has been found that signals composed of many component sinusoids or square waves are perceived as tones having a definite pitch. In certain cases, the auditory features combine to produce a chordlike quality; in others, with closely spaced frequencies, beats or roughness result. Stimuli involving a fundamental frequency and various harmonics are generally perceived as having a single pitch. In addition, in many cases, high frequency components of complex tones are masked by the low frequency components. The perceptual qualities associated with such stimuli can provide useful clues as to the methods of frequency analysis and information processing employed by the auditory system. Often, these phenomena are easier to interpret with tonal stimuli than with noise-like sounds, thus providing useful information about the hearing mechanism applicable to noise stimuli.

Ritsma (1967) used pulsive signals consisting of low and high frequency components to determine which frequencies contributed most to the perceived pitch of such stimuli. He found that the low frequency band tended to dominate pitch perception as long as its amplitude exceeded a minimum absolute level. Many masking studies have shown that low frequencies will mask higher ones more easily than the reverse (Ehmer, 1959a; Ehmer, 1959b; Small, 1959). This suggests that the travelling wave on the basilar membrane (Bekesy, 1960) may interfere with the wave pattern which would normally propagate due to the high frequency components of the stimulus.

Several investigators have attempted to analyze the origin of the perceived pitch of complex tones (Plomp, 1967; Nordmark, 1972). These studies have found that perceived pitch is determined more by the overall periodicity of a stimulus than by its fundamental frequency. This has suggested that in order to understand how complex tones are processed by the auditory system, one must conceive of the ear as a temporal pattern analyzer rather than simply as a mechanical frequency analyzer (Nordmark, 1972). That is, the perceived pitch of a complex tone is, according to Nordmark, related to the time interval between neural pulses.

Because different waveforms and various composite signals are often perceived as having very similar tonal qualities, questions arise concerning what types of confusions listeners might make when asked to discriminate between sounds differing in such dimensions. Webster et al. (1973) investigated this problem using sixteen complex sounds differing along the dimensions of fundamental frequency, waveform, frequency of formants, and number of formants. Each of the four dimensions had two possible values, thus making up the sixteen different sounds. The goal of the investigation was to determine the importance of each of the features in identifying the complex sounds. The findings from this and another related study (Howard and Silverman, 1975) are of fundamental interest in terms of complex tone perception. In addition, the techniques used to analyze confusion patterns are applicable to the noise discrimination studies presented later.

In the experiments conducted by Webster and his colleagues (1973), subjects were trained to identify each dimension of a given complex sound with the value (1 or 2). Thus, these four dimensional sounds were identified by such labels as 1121, 2211, etc. Webster et al. found, as had previous researchers, e.g., Eriksen and Hake (1955), that fewer confusions were made between sounds as the number of dimensions on which they differed increased. Furthermore, sounds differing along the single dimensions of fundamental frequency or waveform were seldom confused with one another. When the single differing dimension between sounds was formant frequency, they were easily confused, and when two sounds differed only in number of formants, they were almost indiscriminable. When the difference between two sounds consisted of the combined dimensions of formant number and frequency, the sounds were also easily confused. Thus, source waveform and fundamental frequency were the dimensions most easily discriminated, with formant parameters being difficult to discriminate. This is somewhat surprising since these formant parameters are believed to carry the major acoustic clues used in vowel identification (Webster et al., 1973).

There were many cases in Webster's investigation in which sounds were not completely identified, but subjects correctly identified three of the four dimensions. With sixteen sounds differing along four dimensions, they contained no redundant information. The authors believe that if the sounds had contained redundancy, more of them would have been completely identified (Webster et al., 1973).

Although an analysis of the errors made by listeners in classifying multi-dimensional stimuli can provide information about feature saliencies, such information can also be obtained through subjects' similarity ratings. Howard and Silverman (1975) asked listeners to provide similarity ratings for all possible pairs of sixteen complex sounds. These were the same sounds used in the experiments reported by Webster et al. Again, sounds differed in fundamental frequency, waveform, and two formant parameters. Howard and Silverman used a multi-dimensional scaling analysis to determine the relative saliency of each feature. The model assumes that "an individual subject's judgment of stimulus similarity is a decreasing linear function of the interstimulus distance" in the perceptual space (Howard and Silverman, 1975).

The results obtained by Howard and Silverman roughly parallel those found by Webster et al. (1973). However, they emphasize the fact that features in the perceptual space are not necessarily analogous to acoustic features. That is, psychological dimensions are not perfectly correlated with the physical dimensions of a stimulus. Analysis of the data indicated that subjects' similarity ratings could be accounted for using three perceptual dimensions. One of these correlated highly with fundamental frequency, one with waveform, and the third with some combination of the two formant features.

Another finding of the Howard and Silverman study was that individual subjects differed greatly in terms of which features they regarded as important when making similarity judgments. That is, different subjects emphasized or de-emphasized various features in the discrimination task. For example, subjects with musical training tended to rely heavily on the dimension of fundamental frequency while ignoring more complex spectral features. In this regard, the feature extraction processes involved in complex sound discrimination may differ greatly between the speech and nonspeech modes. The sounds reported in the previous two studies possessed both tonal and vowel-like qualities, but they were completely meaningless in contrast to familiar speech sounds.

2.3 Relevant Facts from Speech Discrimination Studies

By far the greatest quantity of literature involving complex sound discrimination is in the area of speech recognition. This is especially true when one is concerned with information conveyed by the temporal structure of signals. In fact, the concepts of extraction and processing of auditory features originated with studies of speech recognition (Tartter and Eimas, 1975; Zhukov and Christovitch, 1974; Stevens and House, 1972; Peters, 1967). These investigators have provided conclusive experimental evidence supporting a feature extraction model of auditory perception. For example, it has been shown for a large variety of stimuli that features may be forgotten independently (Stevens and House, 1972).

In the recognition of speech, the feature or perceptual unit of information processed by the auditory system has been hypothesized to involve acoustic elements at the phonemic or subphonemic level (Tartter and Eimas, 1975) or much larger time segments of syllabic length (Massaro, 1972). Most of the information conveyed by speech stimuli concerns the nonstationary or time-varying parts of the signal (Mundie, 1970). That is, the relevant information for recognition is contained in the transient spectra, attack and decay times of various formants, lengths of silent intervals, etc. The complex sounds with which this thesis is concerned include stationary noise bands and amplitude modulation which, although conveying temporal information, is mainly periodic in nature. Thus, the statistics of these temporal shifts may be regarded as stationary if one chooses a sufficiently long time frame.

In the processes of speech recognition, relevant features are extracted mainly from the nonstationary components of the incoming signal, and these features are then compared to patterns stored in memory (Peters, 1967). Unlike the noise patterns to be investigated in this thesis, the internally stored patterns associated with speech recognition are highly overlearned (Stevens and House, 1972). Furthermore, since in addition to recognizing speech sounds, the human must also employ the same auditory features in the production of speech, it is probable that some interaction exists between

production and recognition of speech (Stevens and House, 1972; Peters, 1967). Such interactions must certainly affect the ways in which the information is processed by the auditory system.

As mentioned earlier, it has been shown that discrimination in speech and nonspeech modes involves fundamentally different processes. Massaro (1972) states that the differences may stem from the fact that listeners are extremely familiar with the features of speech sounds, whereas this overlearning does not exist in the discrimination of noise-like sounds. The fact that humans do not possess an interactive mechanism for the production of machine noise may also result in fundamental differences between the two modes. In addition, differences may result because speech-like features are nonstationary, whereas in the present context, the information in the noise sounds of interest is conveyed by stationary properties of the signals.

Since fundamental differences do exist between the two processes, the results of studies on speech discrimination are of questionable value in the understanding of discrimination in the nonspeech mode. The one exception to this is the general concept of feature extraction. The two modes are obviously analyzed by the same peripheral processes (Stevens and House, 1972), and the extraction of acoustic features has been shown to take place in the peripheral stages (Zhukov and Christovitch, 1974; Mundie, 1970). Clearly, the auditory system has no way of distinguishing if a stimulus belongs to a class of speech or nonspeech sounds until some information has been extracted from

the sound. Beyond this however, results obtained with speech stimuli are probably not applicable to problems of noise discrimination.

2.4 Discrimination of Noise-Like Sounds

The types of features which an observer extracts from a complex noise sound in making discriminations are not well understood at the present time. Very little research has been done in this area, and the few studies which have been reported show contradictory evidence as to the type of analysis performed by the auditory system. Some investigators have shown that decisions about complex sounds involving broadband components are made solely on the basis of the most detectable third octave (Fidell et al., 1974). However, another study has shown that the psychological features used by subjects correlate best with an overall spectral shape of the sound (Howard, 1977). These and other studies will be reviewed below as background material for the present work whose goal is to gain an understanding of the information processing capabilities of the human auditory system with respect to noise-like stimuli. Again, the term "noise-like" refers to sounds which can potentially convey information in a form other than speech or music. These include broadband spectra, as well as sounds with various rhythmic qualities. Several investigators have examined the discrimination of sounds in a marine environment, as related to the performance of sonar operators (Janota, 1977; Howard, 1977; Corcoran et al., 1968; Corcoran et al., 1970), and a few studies have looked at problems related to detection and/or identification of engine sounds (Webster et al., 1969; Fidell et al., 1974).

In a fundamental study, Green (1960a) investigated the detectability of a band of noise in noise. Observers were asked to indicate whether a band of noise was present or absent on a given trial. Green found that the human could be modeled as a somewhat inefficient energy detector. That is, the detectability of a band of noise was proportional to the bandwidth, the duration of the signal, and the level of the signal above the background. He also found that the detectability of a band of noise was not affected by changing the center frequency of the band, as long as the signal bandwidth remained wider than a critical band at the given frequency. The results and a brief description of Green's mathematical development will be presented in Chapter III, since the detectability of a given component of a signal will be important in analyzing the discrimination of multi-component signals. Green's results apply only to single bands of noise, and one major purpose of this thesis is to study how these component detectabilities may combine or otherwise be affected in multiple-feature environments.

Janota (1977) has investigated a large class of noise-like sounds using an experimental paradigm called the "Modified Threshold Procedure." These include both marine sounds and laboratory-generated sounds. In his experiments, subjects were presented with two stimuli which differed in the presence or absence of a single "dichotomous" feature. The signals also contained some "fixed" features present in both stimuli. One of the two stimuli was then presented in a background

Gaussian noise, and subjects were asked to identify the sound as either signal "A" or "B." During the response period, signals were initially presented at a very low signal-to-noise ratio, SNR, and this SNR increased with time until subjects were able to make a discrimination decision. The parameters of interest were the SNR necessary to make a terminal decision and the probability of a correct response.

In Janota's studies, dichotomous features included bands of noise as well as amplitude modulation. The modulation studies will be discussed in Section 2.5 of this thesis. Janota found that for a large class of signals involving bands of noise dichotomous, the detection of the dichotomous band was both a necessary and sufficient condition for making a discrimination. That is, in order to reach a discrimination decision, an observer must first have an opportunity to detect the relevant feature. Furthermore, in many cases, the detection of the dichotomous feature resulted in a discrimination decision. The detectabilities of features at the point where subjects responded compared favorably with Green's results for detection of noise bands (Green, 1960a). However, performance was degraded somewhat due to the increased uncertainties resulting from the experimental paradigm used. The result that performance on such discrimination tasks could be predicted based on the detectability of a dichotomous feature was true whether the dichotomous bands of noise originated from high-pass filtering of marine sounds or from deleting a band-limited component from laboratory-generated sounds.

Janota found that discrimination between sounds with dichotomous noise bands could be modeled in terms of feature extraction followed by a hypothesis test on the energy in the dichotomous band. This gave good results for signals when the dichotomous feature was present. He found, however, that in some cases subjects responded at a higher SNR for signals with the feature absent. The model first used by Janota is equivalent to a statement that the listener applies a band-pass filter to the portion of the spectrum containing the dichotomous feature. According to the original model, performance should be independent of the rest of the signal, i.e., no information in the invariant features is used. However, this is not valid in the feature absent case wherein, without the invariant portion of the signal, the subject could never decide that the signal was far enough out of the noise to make a discrimination. That is, looking only at the band containing the dichotomous feature, he would never have a reference to decide that the feature was absent. Janota modified the model in this case to include simultaneous monitoring of the fixed feature band, concluding that it too must be detected in order for subjects to make a "feature absent" decision. In experiments where the invariant feature was approximately 3 dB more detectable than the dichotomous feature, the SNR to respond was nearly the same in both the feature present and feature absent conditions.

The modified threshold procedure used by Janota is the same experimental design used in the experiments to be reported here, and will be discussed in detail in Chapter IV. The decrement in performance

which Janota found in comparing his results with those of Green (1960a) was on the order of 5 dB. Although the amount of the decrement was affected by the type of signals and the probability correct, the difference in performance very closely followed a pattern predicted by Stallard and Leslie (1974). They stated that differences between a task like Green's and a real-world passive sonar detection task should be on the order of 5 dB. These differences are the result of uncertainties in frequency and onset time of the signals, fluctuations in the signals, and effects related to the slowly increasing signal-to-noise ratio. Since the modified threshold procedure involves a sequential decision-making task in which SNR increases with time, a decrement in performance would be expected due to a subject's imperfect memory. This effect corresponds to that observed in classical threshold experiments such as the method of limits.

However, taking these uncertainties into account, Janota does show that the noise discrimination problem can be accurately modeled as one of feature extraction and feature detection. However, the ease with which an observer can extract and detect a given feature may be affected by other components of the signal in more complex discrimination tasks (Norman and Bobrow, 1975). The model developed by Janota is a good starting point for a study of these more complex tasks presented in later chapters.

Several other studies have investigated the salient features in the identification of complex marine sounds as related to the training

of sonar operators (Corcoran et al., 1968; Corcoran et al., 1970). In these investigations, synthetic sounds were produced consisting of whines, roars, and rhythmic components. Corcoran provides evidence that the information content of these signals is closely related to the acoustic features and the signal-to-noise ratio of presentation. Among the conclusions drawn from these studies are the following:

1. Verbal descriptions of the acoustic features comprising the sounds were helpful in training observers to recognize them.

2. The ability of listeners to learn the sounds was dependent on the order of presentation. Changing one relevant feature per item resulted in better learning than did random presentation. Also, learning was facilitated if easy, high SNR sounds were alternated with difficult, low SNR sounds.

3. When information was provided to both the visual and auditory senses, greater emphasis was placed on the visual information when signals were presented at a low SNR.

4. Learning was of a general nature; i.e., subjects learned the sounds according to types of features rather than learning the specific recordings used during training sessions. This finding suggests that information is stored in the form of features, which can be combined to facilitate recognition of new sound patterns. This result, along with that from speech recognition studies showing that features may be forgotten independently (Stevens and House, 1972), lends support to models of complex sound identification involving information stored in the form of general features extracted from the sounds.

Howard (1977) used a multi-dimensional scaling analysis to determine the salient features of eight underwater sounds. This technique, based on subjects' similarity judgments between sound pairs, is the same one reported in an earlier study by Howard and Silverman (1975) involving tonal stimuli. Howard found that the salient features of these passive sonar recordings could be described quite well in terms of a two-dimensional psychological space. The first of these dimensions correlated highly with overall third-octave spectral shape of the sounds. The second dimension was related to a low frequency periodicity present in some of the signals. The relative importance of these two dimensions was found to differ among subjects, suggesting that feature saliency is not determined completely by the acoustic environment, but rather is affected by factors such as prior experience of the listeners.

With the eight signals used, narrowband spectral analysis showed substantial redundancy. The most significant finding of the study is that under these conditions, subjects base their judgments on overall spectral shape rather than on any specific narrowband properties of the signals. Howard states that this finding is "consistent with a multi-level auditory recognition system where early processing involves 1/3-octave or similar spectral analysis, and later processing reduces the spectrum to a more statistically efficient set of psychological features."

Howard's finding that subjects are responsive to overall spectral shape of a complex sound rather than attentive to any specific

narrowband properties is in contradiction to results reported elsewhere (Fidell et al., 1974). Fidell sought to determine the detectabilities of various aircraft sounds in several different noise backgrounds. He found that the detectability of the overall sound was most accurately predicted by an observer's sensitivity to the most detectable feature of the signal. This "most detectable feature" model predicted performance better than a statistical summation model in which component detectabilities were added as orthogonal vectors. Fidell concluded that subjects ignored information in spectral regions which did not contain the most detectable acoustic feature. That is, they used information only in the third octave having the highest signal-to-noise ratio.

The apparent contradictions between the results of these investigators may be due in part to differences in the stimuli used. However, Fidell's experimental technique and analysis of results must be called into question for a number of reasons. First, his article gives no explanation of how component detectabilities were determined other than to state that 1/3-octave analysis was involved. In addition, some of the signals used in Fidell's experiments contained amplitude modulation, and he states, "the presence of modulation caused no special problems." Other investigators have studied amplitude modulated signals and have concluded that amplitude modulation is an important conveyor of temporal information (Janota, 1977). Furthermore, such modulation has an associated detectability, which in many cases

causes it to be the most prominent feature of a signal (Janota, 1977; Dubrevski and Tumarkina, 1967).

The discrimination of noise-like sounds, although poorly understood, arises in a number of practical situations. One such situation involves the need for an observer to distinguish which of several pieces of machinery may be running at a given time. Webster et al. (1969) trained listeners to identify several diesel engine sounds in an attempt to learn which components of the sounds were most relevant to the task. He found that, in general, different engine types were easier to identify than were differences in running speed. In addition, as might be expected, the number of correct identifications was inversely proportional to the amount of masking noise present during the trials. He, like other researchers, also found that providing subjects with feedback about their performance had little if any effect (Gundy, 1961; Robinson and Watson, 1972).

One prominent feature which in many cases can be used to discriminate among noise-like sounds including marine sounds and those of engines and machinery is the temporal structure of the signals. Temporal information may be conveyed by changes in a number of parameters. The final section of this chapter deals with temporal changes in the amplitude of a signal. Low frequency amplitude modulation is generally perceived as giving sounds a rhythmic quality which is often useful in discriminating between them.

2.5 Discrimination of Amplitude-Modulated Sounds

Just as in speech signals, much information about noise-like sounds can be conveyed by their temporal properties. This section deals with the discrimination of temporal changes in the amplitude of such sounds, specifically, amplitude changes which are periodic. The sounds under study involve bands of noise which are amplitude modulated, resulting in a periodic, incremental change in intensity. At low modulating frequencies, below about 20 Hz, these sounds are perceived as having a rhythmic quality. If the modulating frequency is greater than 20 to 40 Hz, the ear can no longer track the periodic intensity changes. The sound then loses its rhythmic quality and is perceived as "wheezing or rough" (Dubrevski and Tumarkina, 1967; Miller and Taylor, 1948). The rhythmic sounds resulting from low frequency amplitude modulation occur in many real-world noise sources including engines, rotating blades, and certain types of cavitation.

The acoustic features which lead to the discrimination of amplitude modulation do not appear in a 1/3-octave or similar spectral analysis. Miller and Taylor (1948) have shown that the spectrum of a noise signal is practically independent of the presence of amplitude modulation. Therefore, the perception of amplitude modulation is based on perceiving the periodic time variation in noise intensity rather than involving the spectrum analysis mechanisms of the auditory system.

Dubrevski and Tumarkina (1967) used modulated white noise to determine the threshold for perceiving amplitude modulation as a function of modulating frequency. They produced curves showing the percent modulation for which the signals were just distinguishable from unmodulated noise for modulating frequencies between 0.5 Hz and 100 Hz. Their data showed a minimum threshold for frequencies around 2-5 Hz, with thresholds increasing for higher and lower frequencies. The performance of subjects was unreliable in the range from 20-40 Hz. This probably reflected the fact that when the sound changed from a rhythmic to a wheezing quality, the criterion used for discrimination must also change.

In investigating the discrimination of amplitude-modulated bands of noise, one pertinent question to consider is: under what conditions can an observer discriminate between a band of noise which is modulated and one which is not? With such signals, the modulation appears as a repeated burst of noise impressed on a continuous noise background. If the modulating frequency is low enough, below about 20 Hz, the sound is perceived as having periodic amplitude changes. Then, the parameter of interest in determining the threshold of discriminability is the ratio of the change in intensity to the average intensity. Weber's law states that the just noticeable difference in intensity is dependent on the intensity of the steady-state signal, but that the ratio ($\Delta I/I$) is a constant (Stevens, 1951). This statement has been shown to be true in a wide range of cases. However, as will be discussed below, the value of the Weber fraction, ($\Delta I/I$), is affected by the duration and the bandwidth of the intensity increment.

Green (1960a), in studying the detectability of noise bands, has determined that detectability increases with signal duration up to about 250 msec. Beyond this point, detectability is relatively independent of signal duration. In the limit for broadband noise modulation where the duration of each noise burst is greater than 250 msec, the Weber fraction is relatively constant having a value given by:

$$\frac{\Delta I}{I_{av}} \approx 0.1 \text{ (Stevens, 1951)}$$
$$10 \log \frac{\Delta I}{I} = -10 \text{ dB.} \qquad (2)$$

Miller and Taylor (Miller, 1948; Miller and Taylor, 1948) investigated the discrimination of short bursts of noise. They periodically interrupted a continuous noise at various rates, producing an effect analogous to amplitude modulation. Among the conclusions drawn by these two studies are the following:

1. Noise interrupted at a steady rate has essentially the same spectrum as does continuous noise.

2. The point at which periodically interrupted noise becomes an indistinguishable series of pulses is a function of the interruption rate and the duty cycle. The critical modulating frequency is about 20 Hz with a 50% duty cycle.

3. For short bursts of noise, the differential threshold for intensity increases as the duration of the added burst of noise decreases.

Moore and Raab (1975) added noise bursts to a continuous noise background to determine the Weber fraction under various conditions. They found the Weber fraction to be a decreasing function of both the duration and the bandwidth of the added increment. This finding is reasonable if one considers the problem from the point of view of an energy detector (Green, 1960a), although Moore and Raab's results are not numerically equivalent to those predicted by an energy detection model. Table 1 gives some of their values for the Weber fraction at various signal durations and bandwidths. The last value in the table is in good agreement with the constant value of the Weber fraction for broadband, long-duration signals. Unfortunately, Moore and Raab's data do not show any simple mathematical relation for calculating other Weber fractions. They do, however, propose an empirical approach to this problem.

In addition to investigating the discrimination of sounds with bands of noise dichotomous, Janota (1977) used the modified threshold procedure to see whether or not the discrimination of amplitude modulated signals could be handled in a similar manner. In these studies, the dichotomous feature was the presence or absence of amplitude modulation on a band of noise. The signals used included marine sounds with quasi-periodic intensity changes between 7 and 10 Hz as well as laboratory-generated sounds which were amplitude modulated with a 10-Hz square wave. Subjects were presented with two signals, one of which was modulated. Their task was then to indicate which signal was presented in the noise during the response period.

TABLE 1

VALUES OF THE WEBER FRACTION AS A FUNCTION OF SIGNAL BANDWIDTH AND DURATION, MEASURED BY MOORE AND RAAB

Bandwidth Hz	Duration msec	Weber Fraction dB
316	10	+5.6
316	250	-3.8
1000	10	-0.9
1000	250	-4.9
3160	10	-1.4
3160	250	-5.5
18000	250	-7.3

Based on the SNR where subjects were able to discriminate the presence of modulation, and the known characteristics of the signals, Janota calculated the Weber fractions for the various features. He found that for the marine sounds and the laboratory-generated sounds, the ratio ($\Delta I/I$) was in good agreement with extrapolated values of Moore and Raab's data (Moore and Raab, 1975). In one experiment where a 10-Hz square wave was used to modulate an octave band of noise centered at 500 Hz, the discrimination threshold was found to be -1.37 dB. For a 10-Hz square wave with 50% duty cycle, the duration of each intensity increment is 50 msec, and the effective bandwidth of the increments was 354 Hz. Using this bandwidth and duration to predict the Weber fraction, good agreement was found with Moore and Raab's data. Since similar results were found with both marine and laboratory-generated sounds of various bandwidths, Janota concluded that the discrimination of amplitude modulation could indeed be modeled as a problem in detection of a dichotomous feature.

In Janota's experiments, when the modulation was absent, subjects responded at a level 6 to 8 dB higher than was found in the feature present case. This suggests that in the case where the feature was absent, subjects needed additional information in order to state that the band of noise was both present and unmodulated.

This chapter has presented a review of findings concerning complex sound discrimination, beginning with complex tonal stimuli and relevant aspects of speech recognition. The Chapter then proceeded

with a more detailed discussion of studies dealing with the discrimination of noise-like sounds including broadband noise and amplitude modulated noise. The following chapter will present discussions of the theory of signal detectability, information theory in the context of auditory information processing, and will conclude with a discussion of a feature extraction model of auditory pattern recognition. These two chapters provide some useful tools for understanding the perceptual processes which underlie the discrimination of noise-like sounds.

CHAPTER III

THEORETICAL APPROACH TO THE DISCRIMINATION TASK

3.1 General

In order for an observer to decide in which of several classes a given sound belongs, he must first perceive the stimulus and then somehow match his perception to his memories of previous stimuli he has perceived. In this chapter, questions concerning how an observer perceives a stimulus, and what information about the stimulus is important for discrimination are investigated. Before reasonable hypotheses about complex sound discrimination can be formed, and meaningful experiments conducted to test these hypotheses, a framework must be established which incorporates known facts about perception and pattern recognition. For the present work, this framework will include discussions of feature detection and information theory as it applies to auditory processing.

The chapter will then provide a basis for assigning numerical detectabilities to the various features of a complex sound. In addition, the discussion will provide some understanding of the acoustic information available to the listener given an environment composed of interacting signal and noise components. The chapter will conclude with the presentation of a generalized feature extraction model of auditory pattern recognition. Hypotheses about how observers

discriminate between noise-like sounds will follow naturally from the theoretical discussions of this chapter. These hypotheses and some experiments designed to test them will constitute the topics of the next chapter.

3.2 The Theory of Signal Detectability

In order to discriminate between two sounds, an observer must at least have an opportunity to detect some characteristic which is different between them. Each feature of a sound has an associated detectability which is determined by the signal and its surrounding environment. So far, the "detectability" of signals and signal features has only been defined intuitively. In this section, a brief mathematical description of signal detection theory will be given. It has been a popular tool in psychoacoustics for over twenty years, and more rigorous treatments of this topic can be found elsewhere (Peterson et al., 1954; Green, 1960a; Green and Swets, 1966; Swets, 1964; Tanner and Sorkin, 1972).

The theory of signal detectability concerns the problem of an observer who, given a stimulus, must decide if that stimulus consists of a signal plus noise or noise alone. In the simplest case, the observer is presented with an input, and he must state, "yes, the signal was contained in the input," or "no, it was not." The observer need not be a human. In fact, the theory was developed to predict optimum performance for an "Ideal Observer." The concept of an ideal observer includes such assumptions as perfect memory and the

ability to maintain stable performance under all probabilities of signal occurrence (Tanner and Sorkin, 1972). Human performance has been found to be suboptimal on such a task, but the theory provides a model with which human performance may be compared.

The theory of signal detectability is based on concepts from statistical decision theory. Given an event X, the observer must decide if X resulted from the distribution of signal plus noise or that of noise alone. Since many distributions of S+N and N alone overlap, almost any input could originate from either distribution, and the listener's decision can only be a probabilistic, statistical one (Robinson and Watson, 1972). In order to decide between these two alternatives, the receiver tests hypotheses about the observation. These are of the form:

H1, the event X is a sample of the S+N distribution.

H0, the event X is a sample of the N distribution.

An example of a decision rule is the likelihood ratio defined as

$$\ell(x) = \frac{P(SN(x))}{P(N(x))} \tag{3}$$

where P(SN(x)) is the probability that event X originated from the S+N distribution, and P(N(x)) is the probability that X originated from the distribution of noise alone (Green, 1960b; Tanner and Sorkin, 1972). An observer using this decision rule would calculate the likelihood ratio and compare it with a number C, called the criterion. Then, if l(x) is greater than C, decide H1, else decide H0. The criterion C

with which the likelihood ratio is compared is a function of such variables as the a priori probability of signal occurrence and the costs and payoffs associated with the various decision alternatives (Green, 1960b; Tanner, 1960; Robinson and Watson, 1972). For example, if H0 is ten times as likely to occur as H1, the observer should set his criterion so as to accept H1 only if l(x) is much greater than in the case where the two alternatives are equally probable. Green (1960b) states that the likelihood ratio or some monotonic transform of it is the optimum decision rule for the following cases:

1. Optimize the expected value of decisions.
2. Minimize risk.
3. Maximize the probability of a correct decision.
4. Set the error rate on some decisicn alternatives at some constant, and maximize the number of correct decisions for the other alternatives.

A topic of primary interest here is the application of the likelihood ratio test to the detection of broadband signals. If a band-limited signal of bandwidth W, arising from a Gaussian process is sampled over a time interval T, then it can be completely represented with no loss of information by 2WT statistically independent samples (Shannon and Weaver, 1949; Green, 1960a). For the detection of this type of signal, the optimum processor is an energy detector (Green, 1960a). Given two independent inputs, one consisting of signal plus noise and the other consisting of noise alone, the optimum decision rule is to state that the input having the larger power most likely contains the signal.

For the likelihood ratio observer, performance is a function of the separation between the means of the S+N and N distributions when the variance has been normalized to unity (Tanner and Sorkin, 1972). This is the detectability index, d'. The detectability is a measure of the amount of overlap between the distributions under H1 and H0 (Robinson and Watson, 1972). In the case where the variances under the two hypotheses are different, an appropriate definition of d' is

$$(d')^2 = \frac{\left[E(x|H_1) - E(x|H_0)\right]^2}{1/2\left[Var(x|H_1) + Var\ (x|H_0)\right]} \quad \text{(Janota, 1977).} \qquad (4)$$

For band-limited Gaussian signals of the type discussed above, this equation gives (Green, 1960a)

$$d'_{opt} = (WT)^{1/2} \frac{\sigma_s^2}{\sigma_n^2 \left[1/2(\sigma_s^2/\sigma_n^2)^2 + (\sigma_s^2/\sigma_n^2) + 1\right]} . \qquad (5)$$

For the case where the ratio of signal to noise power is much less than one, Equation (5) reduces to

$$d'_{opt} = (WT)^{1/2}\ \sigma_s^2/\sigma_n^2\ ,\ \sigma_s^2/\sigma_n^2 << 1 \quad \text{(Green, 1960a).} \qquad (6)$$

As can be seen from these equations, the detectability d' is a monotonically increasing function of the signal-to-noise ratio and is proportional to the square root of both signal bandwidth and duration.

Although the theory of signal detectability was derived for the case of the ideal observer, several investigators have demonstrated its utility in modeling human performance on detection tasks (Green,

1960a; Green, 1960b; Tanner, 1960; Swets, Tanner, and Birdsall, 1961; Robinson and Watson, 1972). At any given probability correct, human performance will be less than optimum due to such factors as signal uncertainties and internal noise. Green (1960a) found that in order for subjects to detect a broadband signal with 75% correct responses, the signal-to-noise ratio must be about 5 dB higher than that required for the ideal observer. Robinson and Watson (1972) state that this type of information loss is a general characteristic of sensory systems.

Another suboptimal characteristic of the human observer is his inability to set a reasonable criterion in cases where the a priori probabilities of the two hypotheses are very different (Robinson and Watson, 1972; Stallard and Leslie, 1974). For example, if the probability of signal occurrence on a given trial is 0.01, and there is some cost associated with a false alarm, the observer will always state that the signal did not occur. However, if the a priori probabilities of the two hypotheses are nearly equal, and the imperfect memory of a real observer is taken into account, the theory does provide a good model for predicting human performance on many detection tasks.

The concepts of detectability for broadband signals can often be applied to broadband features of a signal. It has been shown that when two sounds differ by a single dichotomous feature, the discrimination task in many cases reduces to a hypothesis test on the presence of the feature (Janota, 1977). Component detectabilities

for broadband features can be assigned to the parts of a signal using concepts presented in this section. Similarly, detectabilities can be assigned to features representing amplitude modulation using the Weber fraction discussed in Section 2.5.

However, given a complex environment, the detectability of any feature will be affected by the rest of the signal and noise. The perception of one feature may cue the observer to listen for another, or it may confuse the listener's perception of another feature. In addition to assigning numerical detectabilities to each feature, according to how they would be perceived in a sterile environment, it is necessary to understand the information content of signals composed of interacting components. The next section is concerned with this topic and deals with such questions as the following: How do the signal and noise spectra affect the way in which a feature's detectability is related to the ability of an observer to extract that feature from the stimulus? If two signals differ in several ways, which information is most useful in discriminating between them? When might a feature be relevant, redundant, or confusing? And, in general, to what extent do feature detectabilities relate to discrimination performance?

3.3 Relevant Concepts from Information Theory

3.3.1 General. The concept of information content of signals is rather elusive when dealing with human perception. With respect to an ideal processor, one may discuss the information in a broadband signal in terms of its complete representation by 2WT independent samples.

The human may be treated with some success as a suboptimal processor, with performance on a detection task stated in terms of his efficiency. However, to completely quantify the information processing associated with human perception in this manner is to ignore reality.

Zhukov and Christovitch (1974) have shown that the neural discharge density in the auditory system is such that the neural pathways may be regarded as a continuous, analog information channel. Various researchers have attempted to measure the channel capacity of the auditory system (Oestreicher, 1968; Corliss, 1971). Corliss, for example, used speech intelligibility scores obtained at various signal-to-noise ratios to estimate this quantity. The estimates in the literature, however, differ by several orders of magnitude. Based on data presently available, it is not reasonable to discuss perceptual information content of signals in quantitative terms such as M bits of frequency information and N bits of amplitude information, etc. At best, one can hope to show qualitatively that one type of signal contains more useful information than another.

3.3.2 Information Content of Signal Features. The information which an observer can extract from a broadband signal or feature is, as mentioned earlier, proportional to the signal energy and the signal-to-noise ratio. The noise in the information channel includes that presented in the stimulus as well as internal noise. It has been shown that the internal noise in a detection task is proportional to the external noise (Tanner and Sorkin, 1972; Swets et al., 1959).

With respect to the information available in terms of signal energy, an important parameter is the effective signal duration. It has been found that a listener extracts all useful information from a stationary acoustic signal in the first few hundred milliseconds (Green et al., 1957; Green, 1960a; Tanner and Sorkin, 1972). Therefore, the value of T in equations for signal detectability is generally taken to be 300 to 500 milliseconds (Green, 1960a).

Swets and Green (1961) investigated detection oriented information processing in sequential detection tasks. This type of task, which is representative of many real-world decision-making processes, provides the listener with three choices on a given observation. He can state that the signal was presented, or it was not, or he may defer his decision, awaiting more information. In these experiments, a signal either was or was not presented for a given sequence, and subjects could listen to as many observations as they desired before making a terminal decision. One of the fundamental questions which Swets and Green sought to answer was: Does the observer integrate the information obtained from successive observations in a sequence, or does he treat the observations independently? That is, does the observer make a terminal decision when the combined evidence from all intervals is sufficiently persuasive, or only when the evidence from a single interval is sufficiently persuasive? It was found that subjects do not integrate information in successive observations, but rather they respond when convinced of a decision by the information on a single observation.

However, subjects are capable of integrating information if they are specifically instructed to do so (Swets and Green, 1961). Although the same information was presented on all trials of a sequence, a given observation may convince a subject to reach a decision where an earlier observation did not, because subjects do not maintain a constant criterion across observations.

Having discussed the information content of signals presented in single observation and sequential detection tasks, many questions still remain concerning how the detectability of a signal feature relates to discrimination performance in a complex acoustic environment. In the real world, the discrimination of complex sounds may involve features which are present in one sound and absent in another, or it may involve features which are present in both signals but which differ in their relative detectabilities (Janota, 1977). In the following paragraphs, several studies are presented which address the problem of what useful information can be extracted from certain complex acoustic signals.

In experiments on visual perception, Eriksen and Hake (1955) asked subjects to discriminate between stimuli differing in the dimensions of size, hue, and brightness. Pairs of stimuli differed in one, two, or all three dimensions. They attempted to discern how subjects utilized available information by determining the percent of correct responses as a function of the number of differing dimensions. It was found that when stimuli differed on two dimensions, the discrimination

was more accurate than when either dimension was used singly. The case where stimuli differed along all three dimensions resulted in almost perfect discrimination performance. Furthermore, it was found that performance on the compound-dimension tasks could be predicted by assuming that a subject's judgment about a given dimension was independent of all other stimulus dimensions. The authors caution, however, that this independence may be specific to the task. Some interdependence between features will exist on discrimination tasks, especially in cases where subjects have learned that certain features are correlated and thus expect them to occur in combination.

In general, the concept of independent features implies that an ideal observer could make a discrimination based solely on the most detectable feature characterizing the difference between stimuli. For the ideal observer, differences in other dimensions would be redundant information in a simple stimulus discrimination task. However, Eriksen and Hake showed that discrimination becomes easier for a human observer as the number of features characterizing the difference between stimuli is increased. This suggests that redundant features may convey information which the human subject needs because of his imperfect memory. Some redundant perceptual information is probably valuable in compensating for a human's suboptimal performance on detection and discrimination tasks (Raisbeck, 1963).

Another important question concerns the ease with which an observer can extract featural information from a complex perceptual environment. In what ways does the structure of a stimulus influence a subject's

ability to use information imbedded in the stimulus? Gilliom et al. (1977) addressed this question by presenting subjects with a musical chord, the center tone of which provided a cue to reduce frequency uncertainty on a signal detection task. The task was a yes-no paradigm in which subjects were asked to detect a tone of uncertain frequency. Just prior to each trial they were cued with a musical chord, the center tone of which always matched the frequency of the signal to be detected. Gilliom's hypothesis was, to the extent that an individual component of a chord is tightly grouped, i.e., forms a unified percept with the remaining components of a chord, it will provide a less effective frequency cue. The results showed that:

1. Within a consonant chord structure, performance on the task was poorer when the center component was closer to the high frequency tone than when it was nearer to the low frequency tone.

2. Subjects did worse on the detection task when the information was part of a consonant chord than when it was part of a dissonant one. The featural information was thus more difficult to extract when the signal components were perceptually grouped. In this case, the frequency analysis process was less complete, and thus the cue was less effective.

Stimulus structure has also been shown to have an effect on the feature extraction process in at least one experiment with noise-like sounds (Janota, 1977). As mentioned in Section 2.4, in discriminating between sounds which differ by a single dichotomous feature, subjects

must use information in the fixed feature portion of the signal to decide that the dichotomous feature is absent. The fixed feature must have a detectability higher than that for the dichotomous feature in order that the response SNR be the same for both signal presentations. Janota conducted two experiments of this type in which the dichotomous feature was a low frequency band of noise. In one case the invariant feature was a highly detectable noise band, and in the other it was an amplitude modulated band with the modulation being highly detectable. For both experiments, the response SNR was independent of feature presence or absence in the probe stimulus. However, subjects responded at a higher SNR in the treatment involving amplitude modulation. Janota hypothesized that this effect was due to subjects' inability to attend to the relevant feature because they were distracted by the more obvious modulation component. It has been found that discrimination tasks involving a fixed number of relevant dimensions require more processing time as the number of irrelevant features is increased (Reed, 1973). Thus, perceptual grouping of features as well as distraction by a dominant but irrelevant feature may interfere with a listener's ability to use featural information in a signal.

3.3.3 Theories of Auditory Information Processing. The information processing system which enables humans to discriminate among acoustic patterns has been modeled by a number of researchers (Reed, 1973; Oestreicher, 1968; Massaro, 1972), although its exact representation can only be inferred from experiments designed to measure discrimination

performance. In general, a subject's responses are used to elicit information about decision-making processes. Subjects usually cannot verbalize exactly how one sound differs from another and, hence, they cannot state what decision rules were used in making a discrimination. Nevertheless, experiments concerned with such topics as masking, detection, confusion patterns, and similarity judgments have allowed investigators to describe the information processing system. Most models include the following stages in some form, with the stages representing transformations of the acoustic stimulus.

1. Signal reception and initial encoding into neural pulses--this stage includes the source characterization and actions of the cochlea and auditory nerve.

2. Feature extraction--the features extracted from a signal are not necessarily identical to the acoustic features characterizing the source, but they are correlated with the acoustic features (Reed, 1973; Howard, 1977; Warren et al., 1969).

3. Decision stage--here it is believed that a listener compares the list of signal features with patterns stored in memory. Then, based on the similarity of the stimulus to some remembered pattern, he labels the signal in some manner (Peters, 1967; Reed, 1973).

4. Output or response stage--in this stage, the subject initiates some action based on his earlier decisions. This may include giving a verbal response, pressing one of several buttons, or waiting for more information.

Most descriptions of the pattern recognition system allow for feedback between stages (Oestreicher, 1968; Mundie, 1970). This feedback mechanism allows for compression and funneling of information by allowing a given stage to filter out portions of the signal characterization which are either incorrect or irrelevant. That is, processing at a higher stage may overrule a decision made earlier. Whenever acoustic patterns contain redundant information, the auditory processor can act as an error correcting system (Peters, 1967).

The recognition process will be constrained by the information available in the signal and in memory, the processing resources employed by the subject, and the probabilities of occurrence of various signals (Norman and Bobrow, 1975; Janota, 1977). The available signal information will be a function of the signal duration, bandwidth, and SNR. The processing resources employed by the subject will involve such factors as the effort he expends on the task, the adequacy of his memory, and his familiarity with the task. A listener's ability to recognize a given pattern will be a function of what patterns he expects to hear based on his experience or the instructions he has received. It is doubtful that a subject will correctly identify a sound which has a very low probability of occurrence.

In the first stage of processing, a signal is received by the ear and analyzed by the cochlea. It is well known that the place of maximal stimulation on the basilar membrane is a function of signal frequency (Bekesy, 1960). However, it has been demonstrated in

speech recognition studies that the cochlear transform involves far more than a simple frequency analysis and includes time segmentation of a continuous signal into discrete events (Mundie, 1970; Zhukov and Christovitch, 1974). This idea derives from the fact that the human auditory system is very effective in analyzing nonstationary signals, and characteristics of sounds are available to higher processing stages which would be obscured in a simple frequency analysis. With this scheme, a stationary signal is represented by a sequence of identical time segments. After the cochlea, the information is coded and contained in the varying time intervals between neural pulses in the auditory nerve (Oestreicher, 1968). The pulse density envelope containing the coded information is the result of sampling of the waveform on the basilar membrane (Mundie, 1970).

Massaro (1972) has used masking studies to show that early auditory processing involves a preperceptual unit which:

1. Contains all of the featural information present in the acoustic input.

2. Outlasts the sensory input by as much as 250 msec. Further processing then involves a readout of the information stored in the preperceptual unit. At this point, relevant features may be extracted and the information compressed. Feedback from higher stages may be used in the feature extraction stage to dictate which dimensions are relevant for the particular task (Mundie, 1970). According to Massaro, preperceptual storage is easily interrupted by subsequent auditory

inputs. A second input may interfere with the information in the first unit, thus degrading performance, or the two stimuli may be integrated to form a single pattern.

In the feature extraction stage, decisions are made about various characteristics which identify patterns the listener expects to hear. The acoustic features which characterize the source may include the presence of a band of noise, the presence of modulation, and the frequency or amplitude of a given component. These features are reduced to a set of psychological dimensions which represent the pattern internally. In order to be adequate, a feature list must distinguish between patterns so that no two can have the same description and must be capable of eliminating confusions between patterns which have almost the same set of features (Reed, 1973).

The exact nature of the feature extraction process is not known. However, it has been proposed that the observer performs hypothesis tests on a set of components (Janota, 1977; Reed, 1973). Decisions are then made about the presence of each feature, and estimates of magnitudes are obtained relative to other features. A great deal of debate has arisen as to whether feature extraction and subsequent hypothesis testing are done sequentially or in parallel. Probably, both types of processing are involved, and the extent to which one or the other is used depends on the pattern to be recognized and the parameters of the task (Reed, 1973; Julesz, 1968).

Signal features are not necessarily identified independently (Reed, 1973). When processing several features, the processes may interfere with one another, the first may interfere with the second but not the reverse, or there may be no interference between processes (Norman and Bobrow, 1975). The outcomes of the hypothesis tests may depend on one another as a result of perceptual grouping or distraction by a dominant feature. In addition, in the case of correlated features, the perception of a given component will affect the probabilities of perceiving others. In some cases, a list of features may be inadequate to describe a pattern. Then, structural descriptions must be learned which specify the interrelations among features (Reed, 1973).

In the decision stage, the results of feature hypothesis tests are matched to patterns stored in memory. A subject's decision is then based on pattern similarity (Peters, 1967). When only partial information is known about a stimulus, the observer chooses his response from the subset of stimuli which are consistent with the perceived stimulus information (Reed, 1973). In a simple two-choice discrimination task involving dichotomous features, the presence of the dichotomous feature is often sufficient to classify the pattern. However, in more complex tasks where patterns do not contain dichotomous features, the relative magnitudes of the components must be compared with the memorized patterns. The time and effort required to perform the matching task and make a discrimination decision is a function of the subject's familiarity with the signals. If patterns are highly learned, they will

be more accurately stored in memory, and will be more easily recalled than less familiar patterns.

3.4 A Model of the Discrimination Process for Noise-Like Sounds

Based on the results of previous sections, a model for the discrimination of noise-like sounds will now be developed. The model is based on the detection of dichotomous features in a complex environment in which the ability of an observer to analyze a given feature may or may not depend on the rest of the signal. The proposed model is outlined in Figure 1 for the case of discrimination between two signals S_1 and S_2, which are characterized as follows:

Signal S_1 contains Features ω_1, ω_2, and Feature ω_3.

Signal S_2 contains Feature ω_3 alone.

Thus, the discrimination problem involves two dichotomous features and one fixed feature, present in both signals. In the present context, these features may be bands of noise or amplitude modulation. Some of the noise-like signals to be treated in this thesis involve multiple fixed features, although only one is shown in the figure. To an ideal observer, the fixed feature information is irrelevant when Signal S_1 is presented, although for real observers, the type of fixed feature will be shown to affect the hypothesis tests for the dichotomous features.

The model in Figure 1 consists of the following stages:

1. Signal reception and initial encoding.
2. Feature extraction.

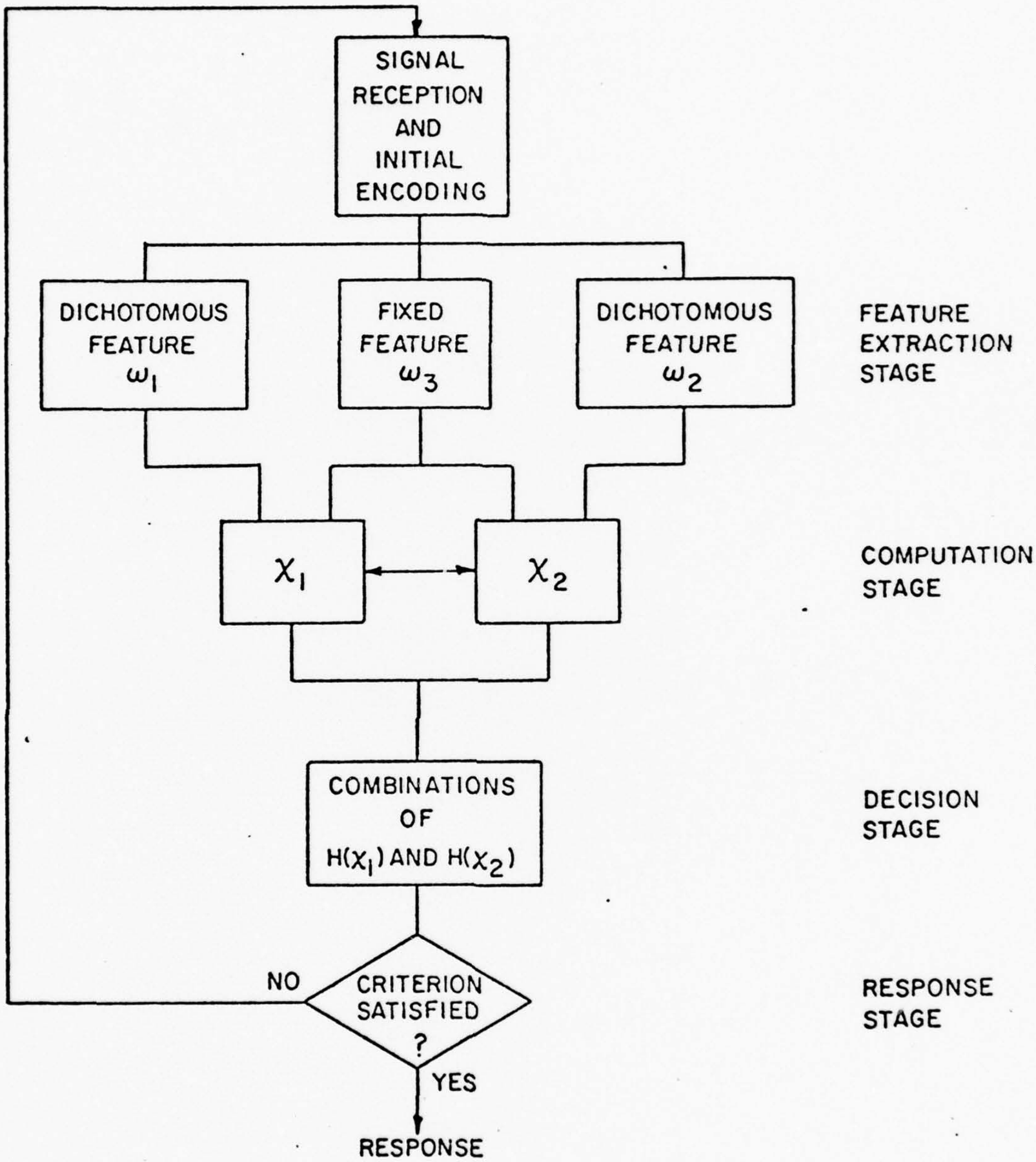


Figure 1. An Information Processing Model of Discrimination Showing Possible Feature Interactions.

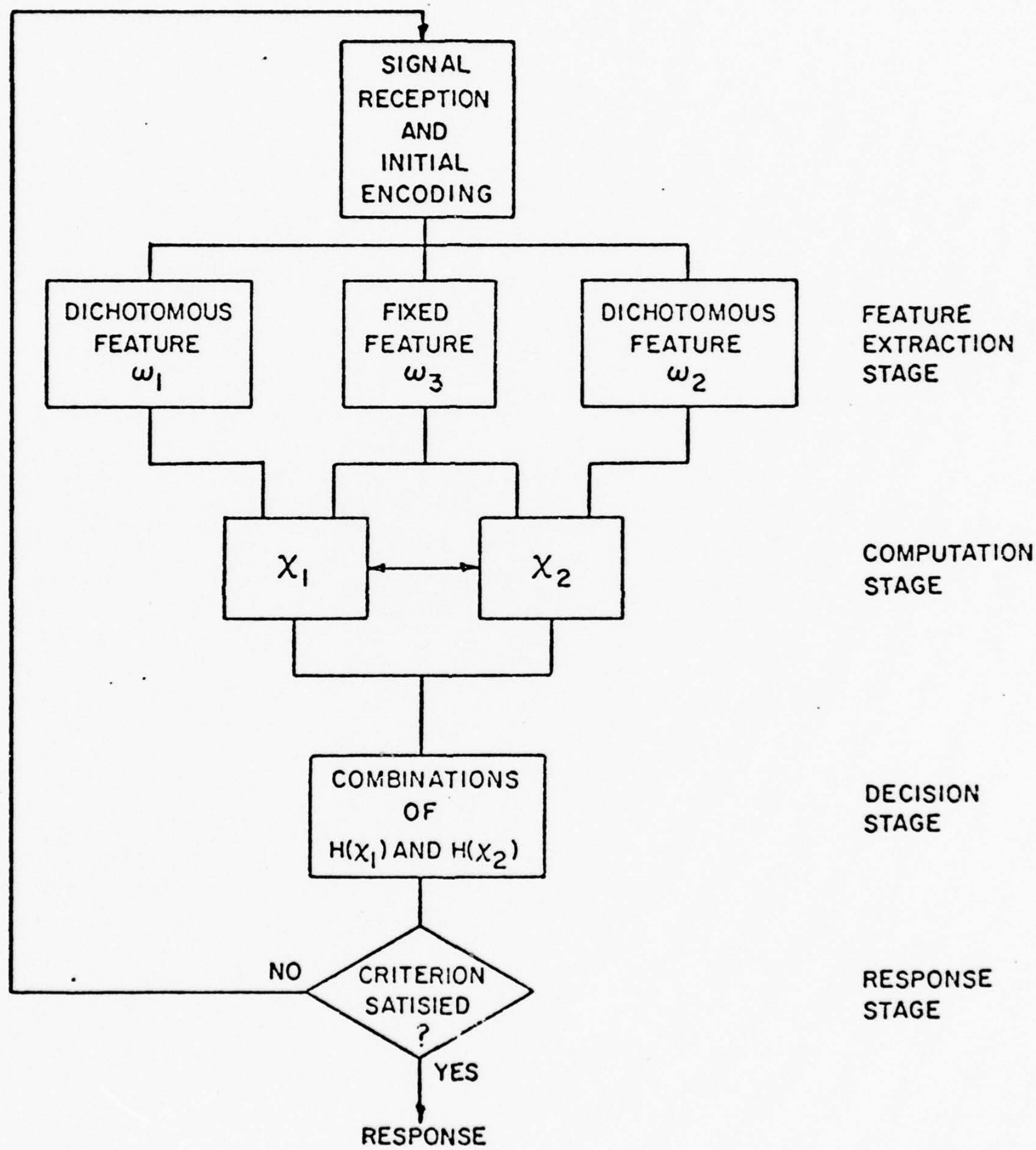


Figure 1. An Information Processing Model of Discrimination Showing Possible Feature Interactions.

3. Computations associated with feature hypothesis tests.

4. Decision stage.

5. Response stage.

The actions which take place in each of these stages will be discussed in subsequent paragraphs. It must be pointed out, however, that this model does not attempt to accurately represent the actions of the human auditory system. It is merely a framework for predicting performance on the types of discrimination tasks illustrated.

The signal reception stage includes the effects of the outer and middle ears, the cochlea, and the peripheral portions of the nervous system. The actions of this stage were discussed in Section 3.3.3, although the details are not relevant to this thesis. It will be assumed that the output from this stage consists of a perceptual unit containing the observables of both the signals and noise (Massaro, 1972; Janota, 1977). Some information compression may occur here, being controlled by feedback from higher centers or from memory (Oestreicher, 1968), although this is not shown in the figure.

The output from the signal reception stage serves as input to the feature extractors. In the model, the process of feature analysis is divided into two parts. The feature extractors here refer to the setting of filter characteristics for noise bands or envelope detection for modulation. Much evidence reported earlier has shown that humans are capable of performing this type of feature extraction. For example, to detect noise bands, filters are adjustable depending on

the nature of the problem. This is true as long as the filter bandwidth is wider than a critical band (Swets et al., 1962). Possible interactions with other parts of the signal are shown as affecting the computations in the next stage, rather than affecting the feature extractors. However, it is possible that the feature extractors themselves are affected by the portion of the signal surrounding them. That is, the adjustment of a given filter may be degraded by other interacting components.

The feature extraction process is shown as taking place in parallel following the pattern of Neisser's model (Reed, 1973). In the present work, this is done merely for convenience, and questions of serial or parallel processing are not crucial to the model.

The next stage of analysis involves computations associated with the feature hypothesis tests. Here, the observer computes some quantity associated with the feature and, in the next stage, compares it to a threshold to determine if the feature is present or absent. If the feature to be tested is a noise band, the computed quantity is related to the signal energy. If the feature involves amplitude modulation, a quantity related to the Weber fraction is computed. The threshold employed by the subject will be a function of his criterion and will include such factors as a priori probabilities, costs, and payoffs, as well as the subject's experience.

It is in the computation stage that interference from other features may affect discrimination performance. Possible interactions

are shown with arrows in Figure 1. The fixed feature, ω_3, may interfere with the detection of the dichotomous features as a result of "information leakage." This information leakage inhibits detection of a dichotomous feature by introducing additional noise into the feature computer. Two possible forms of leakage interference are included in the model. First, degraded performance may result if ω_3 is adjacent in frequency to either ω_1 or ω_2. The adjacency of the fixed feature is thought to cause an overlap of energy because the filter roll-off associated with the dichotomous feature is not infinitely steep. It is hypothesized that the closer the two features are in frequency, the more likely they are to be perceptually grouped. Information leakage from a fixed feature may also degrade the analysis of a dichotomous one if the fixed feature is so dominant as to distract the observer such that he cannot properly attend to the task. This phenomenon was observed by Janota (1977) when certain signals contained a highly detectable but irrelevant feature.

Figure 1 also shows interactions between the hypothesis tests for the dichotomous features. These interactions reflect possible correlations between features, wherein the detection of a given component indicates the presence or absence of another. In the example shown in the figure, features ω_1 and ω_2 always occur together, so that the detection of ω_1 provides conclusive information about the computations associated with feature ω_2.

Interactions other than those shown in the model may affect discrimination performance. In addition, in some acoustic environments feature interactions are nonexistent. One type of interference not shown in the figure involves the background noise. If the noise sounds very similar to a signal feature, subjects may confuse the two, resulting in a false signal classification.

In the fourth stage of the model, the results of feature hypothesis tests are compared with internal representations of patterns the subject expects to hear. If no criterion is exceeded, the subject defers his decision and processing begins again with another signal observation. If the dichotomous features together or separately exceed some criterion, the subject decides that Signal S_1 was present. The subject need not detect both dichotomous features, but in some cases, both will influence the decision. Thus, discrimination decisions will be easier for signals involving multiple dichotomous features than for signals involving either of the features singly (Eriksen and Hake, 1955; Green, 1958). If the fixed feature is detected but neither of the dichotomous features has exceeded its threshold, other decisions must be made before the subject can conclude H_0 that Signal S_2 was present. This case involves complexities which are beyond the scope of this thesis. However, it is believed that in order to decide that the dichotomous features are absent, the subject must not only detect the fixed feature, but also remember the relative levels of the signal components (Janota, 1977). Janota has

shown that the fixed feature must have a higher detectability than that required for the dichotomous features in order to decide that S_2 was present.

In the final stage of the model, all of the feature information and decision rules eventually reduce to a single response action based on the H_0 or H_1 decision in the previous stage. The subject responds according to whatever instructions he has received for indicating the outcome of a classification decision. In the experiments for this thesis, the two signals to be discriminated on a given trial were arbitrarily assigned the letters "A" and "B." Thus, memory plays a part in the response stage as it has in previous parts of the model. For example, a subject might recognize a given sound but fail to remember the appropriate letter designation, or he may have a bias toward a given response. These, along with other possible sources of error, will be handled in a later section of this thesis. The experimental design constitutes the main topic of the next chapter which begins with the presentation of a list of hypotheses to be tested. These hypotheses reflect the interactions between features in a discrimination task and will provide a basis for testing this aspect of the model.

CHAPTER IV

EXPERIMENTS IN NOISE-LIKE SOUND DISCRIMINATION

4.1 General

This chapter presents a discussion of experiments in the discrimination of noise-like sounds designed to test some hypotheses concerning multiple feature interactions. The experiments were conducted over a period of seven months at the Applied Research Laboratory of The Pennsylvania State University using sixteen pairs of laboratory-generated sounds. Five graduate students served as subjects for the studies. The processes of data collection and reduction were automated as much as possible to facilitate accurate analysis of the large volume of data needed for the experiments.

Section 4.2 contains a list of hypotheses about how an observer extracts information leading to a discrimination between sounds. These hypotheses follow naturally from the theoretical developments in the previous chapter and include the effects of multiple dichotomous features, multiple fixed features, and amplitude modulation as either a fixed or dichotomous feature. Next, in Section 4.3, the sound pairs used to test these hypotheses are presented. Multiple fixed and dichotomous features are composed of octave bands of noise at various center frequencies and amplitude modulation of these bands by a 10-Hz square wave. The sound pairs are presented in matrix form with stimulus

complexity increasing along each row and column.

The experimental paradigm used for the discrimination tasks is the modified threshold procedure developed by Janota (1977). This paradigm is discussed in Section 4.4. In the experiments, subjects were presented with two sounds which differed by the presence of one or more dichotomous features, and one of the sounds was then presented in a white noise background. The SNR was initially very low and increased as a function of time until subjects, using the criterion of "reasonably certain," were able to state which sound was presented. During a test session, the same sound pair was used for a group of six events with four such groups comprising a session.

Section 4.5 provides a description of the equipment used to generate and record the sounds, construct the trials, and collect the data. Section 4.6 discusses the selection and training of subjects for the experiments. Subjects included one female and four male graduate students, all but one of whom had participated in previous psychoacoustical experiments using the modified threshold procedure. Finally, Section 4.7 presents methods of data analysis for the experiments. The principal measures of interest with the modified threshold technique are the SNR necessary to reach a terminal decision and the probability of a correct response, P(C). Since these variables are functions of stimulus complexity, and since the experiments were constructed with various features becoming detectable at different levels, the data should result in a qualitative hierarchy of the

information content of features in varied acoustic environments. Data are only analyzed for cases in which the dichotomous feature or features are present in the probe stimulus since, as discussed earlier, analysis of the feature absent case introduced complications which are beyond the scope of this thesis.

4.2 Hypotheses on Multiple Interacting Features

The ways in which the acoustic environment, including the composition of signals and noise, affects the performance of subjects on a discrimination task has been discussed in theoretical terms in Chapter III. This development has led to the formation of several hypotheses which are to be tested in the area of noise-like sound discrimination. In addition to its practical value in industrial or marine settings, the use of noise-like sounds permits one to study aspects of stimulus similarity with the subjective factors of familiarity and meaningfulness removed.

In analyzing the roles played by various signal components in a multiple feature environment, it will be assumed that the detectability of a band of noise is independent of center frequency. The validity of this.assumption was verified by Green (1960a). As an example of this assumption, consider the case of discrimination between two signals S_1 and S_2, where S_1 is composed of two features having equal energy, and S_2 is composed of only one of these features. The assumption is that performance will be the same no matter which feature is dichotomous. This assumption is necessitated by the fact that, in

generating complex signals for the experiments, it was necessary to place similar features at various frequencies depending on the experiment. Accepting this point, the following hypotheses are presented with appropriate methods of testing them to be given in subsequent sections.

1. When basing a discrimination decision on the detection of a dichotomous feature, the discriminability of that feature will be largely determined by the acoustic environment. This environment includes such factors as the bandwidths of features present in both signals, as well as the extent to which the masking noise sounds like the dichotomous feature.

2. When signals differ by two or more dichotomous features, the way in which the component detectabilities combine will be determined by the acoustic environment. However, the discrimination task will always be easier than any of the cases where signals differ by only one of these dichotomous features.

3. If a dichotomous band of noise is adjacent in the frequency spectrum to a fixed feature, the two will be perceptually grouped, and the fixed feature will act as a confusion parameter. Thus, subjects will respond at a higher SNR than in the case where the same fixed feature is nonadjacent. In this second case, the fixed feature, depending on its relative detectability, will act as a cue to the observer that the signal is far enough out of the noise to allow a discrimination decision.

4. When signals involve several dichotomous features, one of which is amplitude modulation, discrimination will be based on the detection of this modulation. That is, the feature AM will have an overriding effect, independent of the other signal features, and the perceptual difference between the sounds will be dominated by the modulation.

5. When two signals involve amplitude modulation as a fixed feature, it will confuse the listeners. The SNR necessary to discriminate between these signals will be higher than in the case where noise bands alone are involved. That is, the dominant but irrelevant feature will be a distraction to the subjects.

6. In general, the data will lend further support to a feature extraction model of complex sound identification in cases where the probe stimulus contains one or more dichotomous features.

4.3 Choice of Noise-Like Sounds

In order to test the hypotheses listed in the previous section and to examine the feature analysis processes which are used in discrimination tasks, experiments were conducted using sixteen sound pairs. These were laboratory-generated sounds composed of octave bands of stationary noise at various frequencies and amplitude modulation of noise bands by a 10-Hz square wave with 50% duty cycle. The sounds were produced using a GR-1390 random noise generator, an HP-3722 noise generator, Spectrum LH-42D and SKL band-pass filters, and several components built at the Applied Research Laboratory

including a two-input mixer, a summing amplifier, and a square wave generator and modulator. When signals were composed of noncontiguous noise bands, separate noise generators were used for each band to insure that the phase among broadband components remained random.

Six of the experiments, denoted Experiments 1 to 6, involved sounds used in an earlier study by Janota (1977). Experiments 7 to 16 involved ten new sound pairs. The acoustic features composing the sounds are described in Table 2. The sounds then involved combinations of features with one or more features dichotomous in a given sound pair. Some of the signals were created so that features had equal spectral levels, while others involved features having equal energy, i.e., narrower bandwidth features had higher relative spectral levels. The signals with amplitude modulation were constructed so that, for the pure signal without background noise, the ratio $\Delta I/I$ was on the order of 0.6 measured in the modulated band. This corresponds to a Weber fraction of approximately -2 dB. The intensity increments were characterized by effective durations of 50 msec and bandwidths corresponding to the 500, 1000, and 4000 Hz octave bands. With these bandwidths, durations, and intensity ratios, the modulation was extremely obvious. The addition of background noise to the experimental trials greatly reduced the ratio of $\Delta I/I$ so that, at the lowest SNR'S used, the modulation could not be perceived. The relative spectral levels and Weber fractions for the features composing each signal will be given in the next chapter. These data will be needed to determine the detectabilities of features at the SNR'S where subjects made terminal decisions.

TABLE 2

DESCRIPTION OF FEATURES COMPOSING THE LABORATORY-GENERATED SOUNDS USED IN THE DISCRIMINATION TASKS

Feature	Description
1	Octave band of stationary noise centered at 500 Hz (band 27)
2	Octave band of stationary noise centered at 4 kHz (band 36)
3	Amplitude modulation by a 10-Hz square wave of Feature 1
4	Amplitude modulation by a 10-Hz square wave of Feature 2
5	Octave band of stationary noise centered at 250 Hz (band 24)
6	Octave band of stationary noise centered at 1 kHz (band 30)
7	Amplitude modulation by a 10-Hz square wave of Feature 6

The sixteen sound pairs are listed in Table 3 and are shown in a matrix form in Figure 2. A comparison of Tables 2 and 3 gives a description of the discrimination tasks. For example, it can be seen that Experiment 3 consists of two sounds containing a 4000-Hz octave band of noise which is amplitude modulated, and the signal representing the "feature present" case also contains an octave band of stationary noise centered at 500 Hz. In the table, the designations S_{H1} and S_{H0} relate to the feature present and absent cases respectively. For a given experiment, the order of signal presentation was randomized so that both signals had an equal probability of being presented first in the exposure set.

Figure 2 is a matrix representation of the experiments conducted and represents a convenient way of showing the data. Essentially, stimulus complexity increases along each row and column. With data shown in this form, results such as SNR to respond or probability correct for a given matrix element may be compared with those for other elements of the matrix to determine the effects on discrimination performance of changing either the fixed or dichotomous features.

The first row and column of the matrix contains Experiments 1, 5, and 6. These are simple discrimination tasks involving a high frequency noise band which is fixed and a low frequency band which is dichotomous. Results from these conceptually simple stimuli will be compared with those for experiments in other rows and columns.

TABLE 3

SUMMARY OF SIGNALS USED IN SIXTEEN DISCRIMINATION EXPERIMENTS WITH EACH SIGNAL DESCRIBED IN TERMS OF ITS FEATURES

Experiment	Features		Dichotomous Features
	S_{H1}	S_{H0}	
1	1:2	2	1
2	1:2:3	1:2	3
3	1:2:4	2:4	1
4	1:2:3	2	1:3
5	1:2	2	1*
6	2:5	2	5
7	1:2:5	1	2:5
8	1:2:5	1:2	5
9	1:2:5	1:5	2
10	1:6	6	1
11	1:3	1	3
12	1:3:6	1:6	3
13	2:5:6	2:5	6
14	2:5:6	6	2:5
15	1:2:3:4	1:2:4	3
16	2:5:6:7	6:7	2:5

*Feature 1 has a detectability 0.25 of that in Experiment 1.

DICHOTOMOUS FEATURES

FIXED FEATURES

DESCRIPTION	1 NOISE BAND	2 NOISE BAND	AMPLITUDE MODULATION	AMPLITUDE MODULATION PLUS NOISE BAND
1 NON ADJACENT BAND	1,1 EXP 1,5,6	1,2 EXP 14	1,3 EXP 11	1,4 EXP 4
1 ADJACENT BAND	2,1 EXP 10	2,2 EXP 7	2,3	2,4
WIDE BANDWIDTH	3,1 EXP 9	3,2	3,3	3,4
2 NON ADJACENT BANDS	4,1 EXP 13	4,2	4,3 EXP 2	4,4
2 BANDS 1 ADJACENT	5,1 EXP 8	5,2	5,3 EXP 12	5,4
AMPLITUDE MODULATION	6,1 EXP 3	6,2 EXP 16	6,3 EXP 15	6,4

Figure 2. Summary in Matrix Form of Discrimination Experiments with Stimulus Complexity Increasing Along Each Row and Column.

The second row of the matrix involves cases where the fixed feature is a band of noise adjacent to the dichotomous feature. For the case of multiple dichotomous features, the entry in this row involves a fixed feature which is adjacent to one of the dichotomous features. A comparison of results for Experiments 1 and 10, Elements (1.1) and (2.1), will show if discrimination performance is affected by making the fixed and dichotomous features adjacent.

A comparison of data for Rows 1 and 3 of the matrix will show if changing the bandwidth of the fixed feature has any effect on discrimination performance. Here, the dichotomous features are not adjacent. A substantial change in discriminability of the dichotomous feature in Experiments 1 and 9 as a result of a bandwidth change in the fixed feature would suggest an interactive effect between features. Simple detection models of discrimination would, in this case, require revision since such models make no assumptions about the roles of fixed features in discrimination tasks.

The fourth row of the matrix introduces experiments involving multiple, nonadjacent fixed features. A comparison of results from experiments in this row with those in Row 1 should determine the effects of additional irrelevant information both for the case where the dichotomous feature is a band of noise and where it is amplitude modulation. A comparison of results between Rows 4 and 5 permits analysis of a still more complex stimulus structure. Row 5 again involves multiple fixed features, but for the experiments in this

row, one of the fixed features is adjacent to the dichotomous feature.

In the sixth row of the matrix, the fixed feature includes amplitude modulation. It was in Experiment 3, Element (6.1), where Janota observed that his subjects were distracted by this highly detectable but irrelevant modulation component.

Each row of the matrix involves a change in fixed features, and data from these experiments will be used to test the information leakage hypotheses discussed in Sections 3.4 and 4.2. In contrast, each column of the matrix involves a change in the dichotomous feature structure of the signals. In Column 1, the dichotomous feature is an octave band of noise. In Column 2, two noise bands are dichotomous. The experiments in the third column involve amplitude modulation as a dichotomous feature. Finally, in Column 4, the dichotomous feature is an amplitude modulated band of noise. If Hypotheses 2 and 5 in Section 4.2 are true, the discrimination tasks in Columns 2 and 3 should be easier than corresponding tasks in Column 1. Column 4 is primarily of interest for analyzing differences in response SNR between the feature present and feature absent cases. This topic will not be discussed in detail, but data will only be analyzed for cases where the dichotomous feature was present in the probe stimulus. However, since data from an earlier study were already available for one entry in this column, it has been included for completeness.

The experiments corresponding to the missing entries in Column 4 were not conducted since only feature present cases will be analyzed

in this thesis. In addition, several other matrix elements are not defined, either because the signals are impossible to create, or because the experiments would provide no new information about the discrimination problem. For example, Element (2.3) would involve amplitude modulation dichotomous with an adjacent fixed feature. A signal defined in this manner is meaningless, since the feature (AM) cannot exist without an associated noise band. The case of amplitude modulation dichotomous with two fixed noise bands, one being adjacent, is handled by Experiment 12, Element (5.3).

On all experimental trials, one of the two signals composing each test was presented in a background noise. This was basically a white noise with frequencies below about 70 Hz filtered out to avoid audio tape saturation (Janota, 1977). The one-third-octave spectrum of this noise is shown in Figure 3. The background noise was produced with a GR-1390 random noise generator, whose level was initially adjusted by means of a GR-1450 stepped attenuator. On each trial, the signal-to-noise ratio was increased slowly, but the loudness of the composite stimulus was maintained constant at 65 phons. This was accomplished with a balanced mixer and a simple automatic loudness control built at the Applied Research Laboratory (Janota, 1977). The balanced mixer was used to change the SNR in approximately 1/2-dB increments from a very low value where discrimination was impossible, to a much higher value where the tasks were comparatively simple.

4.4 The Experimental Design

The modified threshold technique, previously used by Janota (1977), was used to elicit information about performance on the sixteen discrimination tasks discussed in the previous section. An experimental trial consisted of an exposure set and a response period. In the exposure set, the signals were presented without interfering noise, the first signal being designated "Signal A" and the second "Signal B." The designations were purely arbitrary, having nothing to do with the signal characteristics. During the response period, Signal A or B would appear in a noise background with each signal being equally likely to occur.

Following the design of the modified threshold technique, the probe signal was initially presented at a very low SNR, and the signal was brought out of the noise in 1/2-dB steps every two seconds. The changes were very gradual, and there were no transients to indicate a step change. Subjects were therefore unable to report when the steps occurred (Janota, 1977). Starting SNR'S were randomized, being chosen uniformly from a set of values ranging over 4 dB. This was done in an effort to force subjects to respond at an appropriate SNR, rather than after some estimated amount of time. Unfortunately, as will be discussed later, this effort was not entirely successful. Time-dependent factors which influence criterion cannot be completely separated from factors which are SNR-dependent using this paradigm.

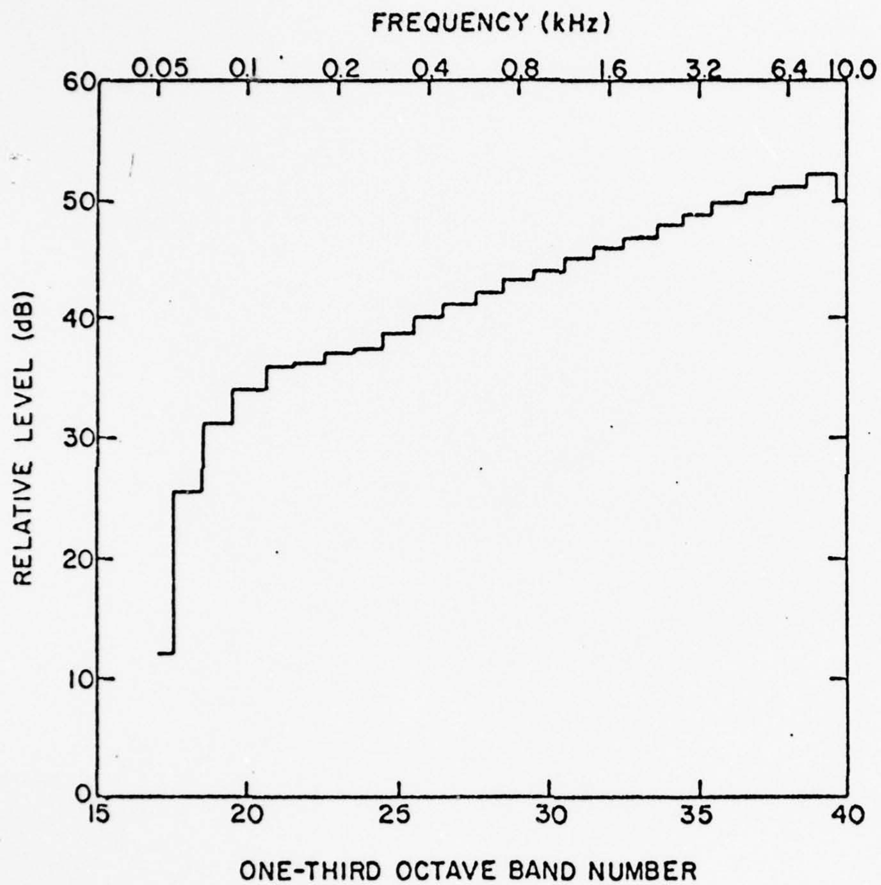


Figure 3. One-Third Octave Spectrum for Background Noise Against Which Probe Stimuli Were Presented.

At the beginning of a response period, the probe signal was completely masked by the background noise. As the trial progressed, the SNR would increase either until the subject was able to respond or until some final cutoff SNR was reached. Subjects responded by pressing one of two buttons marked "A" and "B" on a response recorder. Their answers were recorded with tones on a cassette. After a response, the signal was blanked until the next event, so that subjects received no feedback from hearing the probe at a high SNR. In fact, no feedback of any kind was used during the experiments for the following reasons: First, although the question is still under study, it has been demonstrated that feedback is not effective in improving performance, and it may degrade performance by causing a subject to erroneously shift his response criterion (Gundy, 1961; Robinson and Watson, 1972). Second, the methods of data reduction employed prevented immediate knowledge of results.

Trials were recorded on a Crown 700 tape recorder with twenty-four trials per session. Subjects could conduct tests at their own convenience with the restrictions that no two sessions could run sequentially, and not more than two sessions could be run on the same day. Subjects performed the tasks in an audiometric booth, listening to the tapes over calibrated TDH-39 headphones. The booth was small but comfortable, containing a chair, a ledge to write on, a window, and the response recorder. The Crown 700 recorder on which subjects played the session tapes as well as a cassette machine for recording

response data were located just outside the booth. A subject would mount the assigned tape, insert a response cassette, enter the booth, and begin the session. Sessions lasted from 45 to 55 minutes.

The experimental trials were recorded in blocks of six events with four different sound pairs constituting a twenty-four event session. The four experiments chosen to comprise each session were a mixture of "easy" and "hard" tasks. Research has shown that this type of design increases motivation and improves performance (Corcoran et al., 1968; Robinson and Watson, 1972). During a group of six events, the exposure order for signals remained the same. However, both exposure orders were used for each experiment across sessions. In one session, signals for a given experiment may be presented in the order S_{H0}, S_{H1}, and in a later session, the exposure order would be reversed. On the first event of a group, the exposure set was presented twice followed by the response period containing the signal plus noise. On subsequent events of a group, the exposure set was only presented once prior to the response period.

A total of nine session tapes were constructed with four sound pairs per session, and most subjects listened to each tape twice during the investigation. Across the nine tapes, trial blocks for Experiments 7 to 16 occurred three times each, and blocks of trials for Experiments 1 to 6 occurred once each. Additional data collected previously with the same subjects were available for Experiments 1 to 6. In addition to the nine new tapes used in this investigation,

subjects again listened to one tape made earlier containing Experiments 1, 2, and 3. The repetition of these first six sound pairs during the present study, as well as the fact that subjects listened to each tape twice, permits checks of performance reliability over time. The total number of valid events collected for each experiment will be listed in the next chapter.

The use of the modified threshold technique in these discrimination experiments affords a number of advantages in data collection and interpretation. The sequential nature of this paradigm is reflective of many real-world situations in which more information about a prospective decision can be gained as a function of time. With this method, signals may be presented at many different SNR'S in a relatively short period of time. Classically, detection and/or discrimination trials are presented at many fixed SNR'S, with the result that thousands of data points are needed to analyze subject performance. However, this aspect of the modified threshold procedure, which is an advantage in data collection, proves somewhat troublesome in interpreting the results. This is because the experiments only provide response information at the SNR where a subject is willing to make a high confidence terminal decision. Thus, the technique only provides data about one point on the psychometric function.

Probably the most undesirable aspect of the modified threshold technique is the interdependence of response SNR and response time. As noted earlier, the starting values of the SNR were randomized in

an effort to prevent subjects from responding after a given elapsed time. However, because the SNR is an increasing function of time, some correlation between these variables is inevitable. In fact, this correlation is always very high, and is given by the equation

$$\rho = \frac{\sigma_T}{\sqrt{\sigma_T^2 + 26.67}} \tag{7}$$

where σ_T is the standard deviation of response time. This equation was derived by assuming that SNR increases linearly at 1/4 dB per second, but this is not very different from the 1/2 dB per two seconds stepped increase which was used in the experiments. The derivation of this correlation relation is given in Appendix B. It is shown there that in order to reduce the correlation coefficient below about 0.5, some experimental variables must be changed so as to exceed reasonable limits.

The high correlation between SNR and time is disturbing because it means that criteria based on these two variables cannot be separated. The detectability of features is related to the SNR, making this a quantity for which accurate measures are desired. However, using the modified threshold technique, the response SNR is affected by time-dependent criterion changes.

The trials were terminated at various high SNR values whether or not a subject had responded, and some subjects were observed to become very frustrated if they failed to respond before the trial was

cut off. Thus, it appears that they were assigning a cost to deferring their decisions beyond a given time. Since subjects were never given feedback as to their performance, it is implied that after some time the penalty for not responding was subjectively greater than the penalty for being wrong. However, for certain very difficult experiments to be reported in the next chapter, subjects did not respond on a large number of events. The subjects used in the present studies seemed much more willing to allow an event to terminate without responding than were earlier subjects used by Janota (1977). Therefore, although some shift in criterion almost certainly occurs with time, this effect is not nearly so great as was first believed.

4.5 Equipment

This section briefly discusses the equipment used for recording the experimental trials, playing these trials back and recording responses, as well as the equipment used in the preliminary phases of data reduction. The sixteen sound pairs and the background noise were created using the equipment noted in Section 4.3. Both signals and noise were carefully calibrated to have the desired spectra, and their 1/3-octave spectra were tabulated relative to an arbitrary reference. These signals then served as inputs to a rather intricate recording process designed by Janota (1977).

After the signals for a given experiment were created using the noise generators, appropriate band-pass filters and mixers, the

experimental trials were recorded on "primary tapes." These one-inch, fourteen-channel tapes contained all of the cuts for a given experiment, and sections of the primary tapes were re-recorded onto quarter-inch audio tapes along with verbal instructions to the subjects.

Three channels of a primary tape were used for recording each trial. The first, the signal channel, contained the exposure set and the response period. These signal channels were FM recorded at 60 inches per second. A 12.5-kHz tone was recorded on another channel of the primary tapes during each event. This tone was used to control a phase-locked loop which, in turn, controlled the switching of inputs in the recording of the audio tapes. This facilitated the recording of verbal instructions on the audio tapes as well as eliminating the FM discriminator noise when primary tapes were stopped. A third channel of the primary tapes contained a tone whose frequency was proportional to the SNR on the signal channel. This ramped tone was used later in the automatic data reduction phase of the experiments.

A sequencer designed by Janota controlled the recording of signals on the primary tapes. The sequencer stepped automatically through the two exposure signals and then enabled recording of the signal plus noise for the response period. A balanced mixer was controlled manually to ramp the SNR on each trial, and an automatic loudness control maintained a constant level of 65 phons to insure that subjects incurred no risk of hearing damage.

The audio tapes to which the subjects listened were AM recorded from the primary tapes on a Crown 700 recorder. These were two-channel tapes, with one channel containing the trials and instructions and the other channel containing the ramped tone proportional to SNR as well as additional control signals. Again, recording levels were carefully controlled to insure that subjects were never exposed to excessive sound levels.

Response data were recorded on cassettes with tones of different frequencies corresponding to the A and B responses. The ramped tone, proportional to the SNR, was recorded from the audio tape onto the cassette until the point where a response was made. Thus, for a given trial, the data cassette contained a tone of increasing frequency followed by a response tone. These cassettes were then played into a digital counter, and the tone frequencies were recorded onto digital tape using a Pertec phase-encoded tape drive. The digital tapes served as input to a software package developed by the author. The data reduction software performed the following functions:

1. Determined the response and final SNR for each event,

2. Matched the experimental data with the appropriate primary tape data,

3. Removed from the data set any events which contained errors resulting from a number of possible hardware malfunctions,

4. Tabulated the data in a form compatible with an existing statistical analysis package.

In addition to the equipment used for signal recording and data reduction, the methods used for determining the Weber fractions for modulated waveforms warrant further discussion. The presence of amplitude modulation does not affect the 1/3-octave spectrum of a signal. Rather, information about the peaks and troughs of the modulation must be extracted from the signal envelope. This was accomplished through the use of equipment designed by the author at the Applied Research Laboratory. The fundamental components of this design are shown in a block diagram in Figure 4. The basic principle of this system is that the voltage of the modulation envelope is sampled at the peaks and troughs of the waveform; these voltages are converted to frequencies on a linear scale using a voltage-controlled oscillator, and the frequencies are determined using a digital counter.

The broadband signal was initially modulated with a 10-Hz square wave. This modulated signal was fed to an RMS envelope detector whose output was connected to the analog input of an SHM-1 sample-and-hold device. The analog output of the sample-and-hold circuit was fed through an amplifier to a voltage-controlled oscillator producing a tone whose frequency was proportional to the voltage of the modulation envelope. The "hold" command was controlled in the following manner. A phase-locked loop synchronized with the 10-Hz square wave was used to trigger a monostable multivibrator on either the leading or trailing edge, depending on the position of a switch. The output of this device was then used to set a D latch which, in turn, initiated the

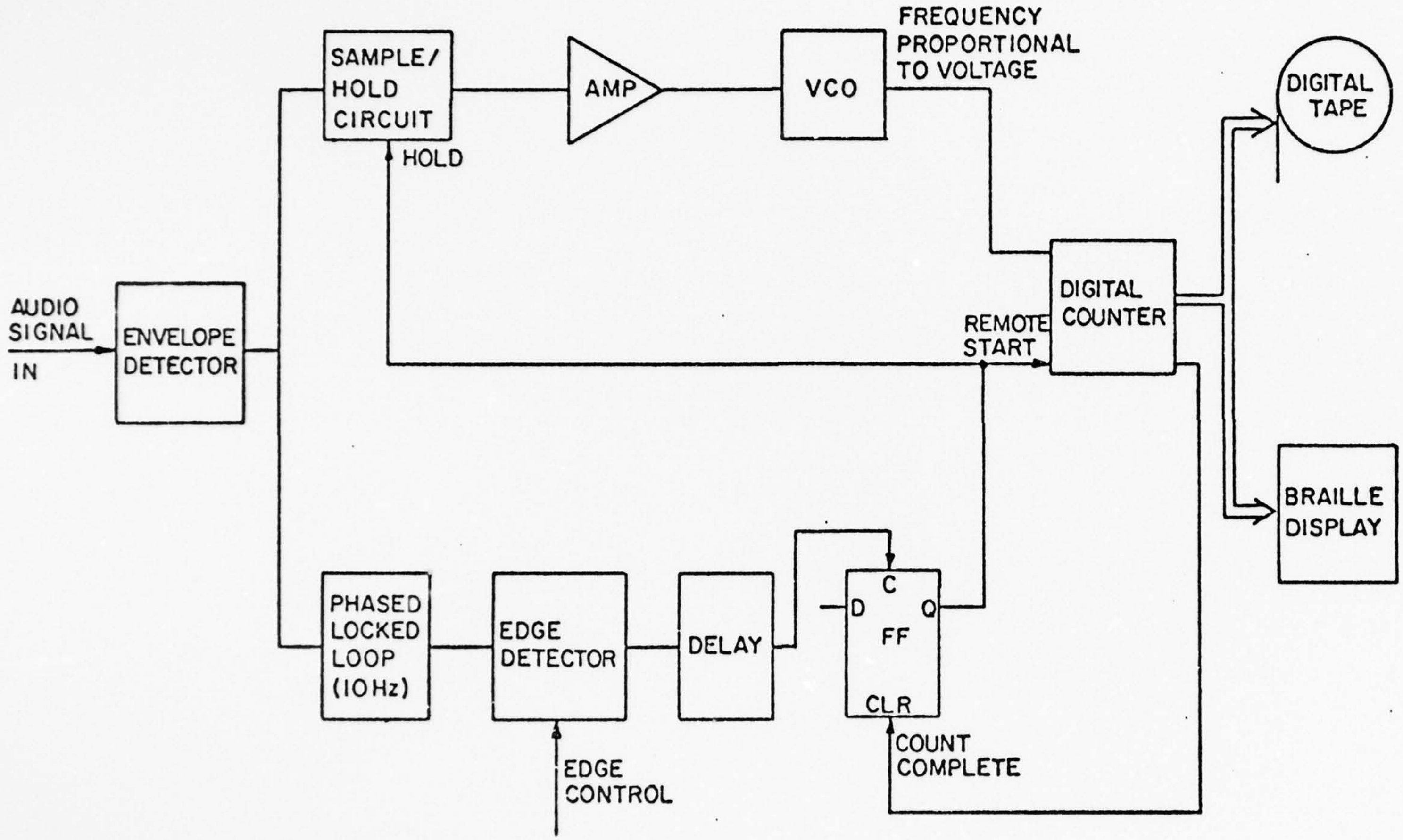


Figure 4. Block Diagram of Equipment Used to Measure Modulation Levels for Laboratory-Generated Signals.

"hold" command to the SHM-1. The pulse time for the monostable multivibrator was set to correspond to the time constants in the envelope detector filter. Thus, the hold command was initiated at the midpoint of either peak or trough of the modulation envelope. The sample-and-hold device no longer tracked the signal, but rather stored this peak or trough value. This voltage was converted to a corresponding frequency, and the digital counter was started at the same time the hold command was initiated. When the counting sequence was completed, a return pulse from the counter served to restart the cycle.

The levels of the spectra and amplitude modulation were checked periodically throughout each recording session. If any calibration error greater than 1/2 dB was found in the spectrum, the primary tape was re-recorded. Much care was taken to insure that all factors associated with the signals and noise remained constant throughout a session, so that subjects' responses on the discrimination tasks were not affected by changing loudness levels or other recording errors.

4.6 Subjects

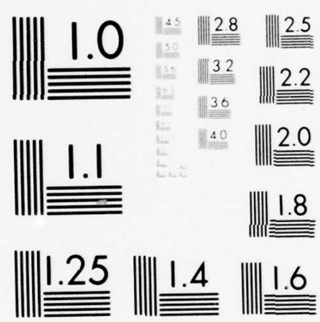
Five subjects participated in the present studies which lasted over a period of about seven months. These were one female and four male graduate students, all but one of whom had participated in experiments using the modified threshold technique just prior to the present work. These earlier experiments included Sound Pairs 1-6 as well as other sound pairs discussed by Janota (1977). Data collected

AD-A057 972   PENNSYLVANIA STATE UNIV UNIVERSITY PARK APPLIED RESE--ETC F/G 6/10
DISCRIMINATION OF NOISE-LIKE SOUNDS INVOLVING MULTIPLE INTERACT--ETC(U)
MAY 78 D W MARTIN   N00017-73-C-1418

UNCLASSIFIED   TM-78-171   NL

2 OF 2
AD A057972

END
DATE FILMED
10-78
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

with Sound Pairs 1-6 in these earlier investigations will be reported in this thesis and, as mentioned earlier, the occurrence of these sounds on the new tapes permits an analysis of reliability and practice effects.

All five subjects were shown to have normal hearing as measured by current audiograms. During the study, one subject had a minor ear infection lasting about one week, and he did not listen to any tapes during this time. At the beginning of each session, the subjects received recorded instructions which included a statement to the effect that no sound would be so loud as to cause discomfort or hearing damage. In addition, the instructions contained information about trial and block structure, procedures for responding, and probabilities of signal occurrence. The precise instructions which the subjects received are given in Appendix A.

In an independent study with a naive subject, the author showed that training in the use of the modified threshold procedure required approximately five sessions for performance to stabilize. During the first five tapes, subject performance improved greatly and then leveled off. No improvement was found aften ten sessions. Therefore, regarding subject training for the present experiments, the first five sessions for each subject will not be analyzed with the performance data since these sessions were required for them to obtain stability. The first five sessions have been disregarded for the subject who was initially naive in this investigation. In addition,

since data collected with the other four subjects in the previous investigation is being used, the first five of these earlier sessions have been omitted for each subject.

Many problems exist when attempting to evaluate subject training for specific signals. In most cases, when a given sound pair occurs to which a subject was exposed in an earlier session, he does not recognize it as one which he has already heard. Descriptions of the complex sounds cannot be easily verbalized or remembered over a period of several sessions. However, subjects trained in the use of the modified threshold technique show very high consistency when data collected for a given sound pair are compared between early and late sessions (Cornell, 1978). In the present study, differences in mean SNR for a given signal pair presented in several sessions are statistically significant in only a few cases. These cases where practice effects are observed will be noted in the data.

4.7 Methods of Data Analysis

So far, this chapter has presented various aspects of experiments conducted to analyze the role of feature interactions in complex discrimination tasks. This section discusses the methods of analysis used to draw meaningful conclusions from the raw data resulting from each experimental trial. This data analysis involves a number of decisions concerning which data can be pooled, which data should be deleted from subsequent analysis, and which statistics of the data will provide measures which are both reliable and valid. In some

cases, these decisions have been based on previous experience with the modified threshold technique, while in others, they are the result of statistical tests showing that some data are more meaningful than others. For example, some subjects indicated that a sample of the signals in noise would have been helpful prior to the first event of a group, and that they used the first event of some groups as such a sample. Subsequently, statistical tests revealed that performance often differed between the first event and later events of a group. Furthermore, this phenomenon had been observed in earlier modified threshold experiments (Janota, 1977). Therefore, in the present data the first event of each six-event group was omitted from the analysis of results. Other modifications of the data set will be discussed after presentation of the performance measures available in the modified threshold technique and the statistics used to interpret these measures.

The three quantities measured using this paradigm are the classification decision, S_{H0} or S_{H1}, on each event, the signal-to-noise ratio at which a subject is willing to make a terminal classification decision, and the time from the beginning of the event until the response. When these data are appropriately grouped, relevant statistics include the observed probability of a correct classification decision and the mean and standard deviation of the SNR to respond. A feature detectability, d'_{mt} can then be computed using the mean SNR and the feature bandwidth. (The subscript "mt" refers simply to

observed detectability using the modified threshold procedure). Detectabilities will be computed using Equation (5), with d'_{mt} substituted for d'_{opt}. Feature detectabilities and values of P(C) for each experiment can then be tabulated in the form of the matrix shown in Figure 2. As discussed in Section 4.3, it is then possible to analyze trends in the data and to draw inferences about how feature interactions affect the detectabilities of the dichotomous dimensions.

On each experimental trial, subjects were asked to make a classification decision about which signal was presented in the noise. These decisions taken over a large number of trials lead to an observed probability of a correct response, P(C), for the experiment. Assuming that events comprise Bernouli trials with equal probability of occurrence, an approximately Gaussian distribution can be obtained for X observed correct classifications in N trials (Janota, 1977). Using the appropriate transformation, the 90% confidence limits on P(C) are given by

$$\{\mathrm{Sin}\left[\left(2\mathrm{ArcSin}\sqrt{\frac{X}{N}}\right) - \frac{1.69}{2\sqrt{N}}\right]\}^2 \leq P(C) \leq$$

$$\{\mathrm{Sin}\left[\mathrm{ArcSin}\sqrt{\frac{X}{N}} + \frac{1.69}{2\sqrt{N}}\right]\}^2 \quad \text{(Janota, 1977).} \qquad (8)$$

A large number of events are needed to make this confidence interval sufficiently narrow to allow tests of statistical significance on the parameter P(C). Fifty to seventy events are required to express P(C) within +-10% of the actual value.

Another performance measure in these experiments is the SNR where a subject is willing to make a terminal decision. This SNR is one of the outputs from the data reduction software and measures the level of the signal above the noise at the point where a response was made. The SNR is here defined to be the average level in dB of the signal relative the noise in the dichotomous band. The SNR is calculated as the sum of 1/3-octave band levels in the dichotomous portion of the signal minus the sum of 1/3-octave noise levels in the same bands. The 1/3-octave spectral plots for some of the signals relative the background at the point of response will be illustrated in Chapter V.

The SNR for each dichotomous feature is given by the equation

$$SNR = \ell_o + \ell_b \tag{9}$$

where ℓ_o is the average level of the feature above the noise at 0 dB balanced mixer setting, and ℓ_b is the balanced mixer setting corresponding to the point of response and corrected for some non-linearity in the mixer. Having thus determined the SNR for each dichotomous feature on each trial, data from similar populations may be grouped to obtain the mean SNR as well as the sample standard deviation. Janota (1977) has demonstrated that the distribution of response SNR'S may be regarded as Gaussian, given a small number of no-response events. Then, the 90% confidence limits on the mean are given by

$$\left[\overline{X} - T(\tfrac{\alpha}{2}, N-1)\frac{\sigma}{\sqrt{N}}\right] < \mu \leq \left[\overline{X} + T(\tfrac{\alpha}{2}, N-1)\frac{\sigma}{\sqrt{N}}\right] \quad , \tag{10}$$

(Freund, 1971)

where $\alpha/2$ denotes the confidence interval for the T-test.

The third performance measure, the response time, is unfortunately very highly correlated with response SNR. Although response time would be an important quantity for determining memory effects and criterion shifts, it cannot be analyzed independent of response SNR as was discussed in Section 4.4.

Given the mean SNR and P(C) for each set of grouped data, the next step in the analysis of results is the computation of detectabilities for the dichotomous features using Equation (5). These detectability calculations are necessary in order that data for features having different bandwidths and spectral levels may be normalized to some common base. A simple comparison of SNR'S among experiments would be meaningless since not all signals were constructed in the same manner. The common basis for comparison among experiments with noise bands dichotomous is then a quantity related to feature energy. For cases with amplitude modulation dichotomous, the ratio $\Delta I/I$ will be calculated at the point of response.

As noted by Janota (1977), a problem arises with the modified threshold technique in determining the integration time, T, associated with the detection opportunity. In the calculation of d'_{mt}, an integration time of 500 msec will be used in this thesis. However, since the purpose of calculating detectabilities is to normalize the data and not to compare the magnitudes of d' with those observed by other researchers, the choice of T in Equation (5) is arbitrary. That

is, the parameter T appears as a multiplicative factor in the equation, such that a change in T would only change the absolute magnitudes of d'_{mt} while keeping proportions constant between feature detectabilities.

In order to perform the above calculations for mean SNR, detectability, and probability correct, decisions must be made as to which data may be pooled in the analysis. Prior to further analysis, the following data have been omitted from processing:

1. All data for Experiment 7. This experiment has been eliminated from consideration because results were so unreliable as to be worthless in providing new information about discrimination performance. The primary tape for this experiment was constructed with the initial SNR so low that the task was nearly impossible. Subjects failed to respond on 40% of the trials, and performance was significantly less than chance on those trials where responses were given.

2. All data for which S_{HO} was the probe. For reasons discussed earlier, data will only be analyzed for cases in which the probe stimulus contained the dichotomous feature.

3. Data associated with the first five sessions for each subject. These tapes have been considered as training for the subjects in the use of the modified threshold procedure.

4. Event one of each six-event group. Many subjects used this first event as a sample of the signal in noise. Earlier investigations have shown significant performance differences between this event and the subsequent five events of a group.

5. All data for which subjects failed to respond. Only in Experiment 7 was the number of no-response events large enough to cause concern. In all other cases, omission of these events from the data set does not result in a significant shift in the response distribution. Subjects failed to respond on less than 10% of the events in each of the fifteen experiments to be reported.

After these omissions, the T-test for difference of means was used to determine which subsets of the data could be pooled in the final analysis. Among the various subsets considered in each experiment were the following:

1. Effect of exposure order. Given that S_{H1} was the probe, significant differences were occasionally found between groups of data having different exposure orders. These differences were not always in the same sense, and a satisfactory explanation for them has not yet been found. However, the decision was made to pool data across these differences because such pooling did not increase the standard deviation by more than 1/2 dB in any case. Furthermore, when significant differences were found, one of the samples usually contained fewer than ten events, thus reducing the power of the T-test as an interpretive tool.

2. Practice effects. As noted earlier, most subjects listened to each tape twice. In two cases, Experiments 8 and 12, significant differences were found in mean SNR when data from these two groups were compared. In both cases, the mean response SNR was lower the

second time subjects listened to the tapes, probably reflecting an improvement in performance with practice. In these cases, only the second set of data has been used for further analysis.

3. Effects of performance differences between individual subjects. Cornell (1978) has investigated the performance variability of trained subjects using the modified threshold technique. He found that between-subject variability was comparable to within-subject variability. In the present experiments, statistically significant differences in response SNR were found between individual subjects in a few cases. Generally however, tests for difference of means were not significant at the 90% confidence level. Therefore, data for all five subjects has been pooled in the analysis. It should be noted that the observed variance in response SNR can be reduced slightly by considering the data for individual subjects. However, these reductions in variance are probably not meaningful in view of the overall goal of this thesis. In addition, the pooling of data across subjects results in a large enough data set to allow fairly accurate prediction of the probability of correct responses.

CHAPTER V

RESULTS AND DISCUSSION

5.1 General

In this chapter, the results of the sixteen discrimination experiments are presented and analyzed in terms of the pattern recognition model discussed in Section 3.4. The data will be analyzed according to the methods of Section 4.7, and the detectabilities and/or Weber fractions associated with the dichotomous features will be reported. It will be shown that some of the hypotheses in this thesis are supported by the data, while others are not. Unfortunately, it was found that some of the results do not adequately address the questions they were designed to answer. However, substantial new information about noise-like sound discrimination is provided, and areas which warrant further study are identified.

Table 4 summarizes data for the measured quantities in the sixteen experiments. Column 3 lists the number of valid events for each experiment after omission of some data according to the criteria listed in Section 4.7. Column 4 gives the mean SNR, Column 5 the sample standard deviation of SNR, and Column 6, the observed probability of correct responses. Again, the SNR refers to the level of the dichotomous feature above the noise at the point where the terminal decisions were made. For Experiments 2, 4, 11, 12, and 15, in which

TABLE 4

SUMMARY OF EXPERIMENTAL RESULTS AND FIRST-ORDER STATISTICS FOR SIXTEEN DISCRIMINATION TASKS INVOLVING DICHOTOMOUS FEATURES

Experiment	Feature	N	$\overline{SNR}$ (dB)	S_{SNR} (dB)	P(C)
1	1	48	7.7	3.24	0.854
2	3	108	2.33	2.49	1.00
3	1	62	9.5	3.38	0.8226
4	1:3	58	3.77	2.33	0.983
5	1	50	6.86	3.07	0.840
6	5	34	5.33	3.11	0.971
8	5	26	7.14	3.24	0.846
9	2	70	4.24	3.64	0.614
10	1	42	7.01	3.05	0.905
11	3	55	-0.84	2.95	1.00
12	3	22	-0.53	1.84	1.00
13	6	57	5.09	3.38	0.491
14	2	50	1.15	3.15	0.960
14	5	50	6.43	3.15	0.960
15	3	53	14.13	5.41	0.906
16	2	44	6.21	2.52	0.932
16	5	44	11.4	2.52	0.932

in which the dichotomous feature is amplitude modulation, the SNR refers to the spectral level of the signal relative the noise in the modulated band. Experiments 14 and 16 each occupy two rows of the table since these experiments involve two dichotomous features having different spectral levels. Finally, no data are shown for Experiment 7 because problems in the tape construction for this experiment resulted in data which do not validly represent the discrimination task. The rest of this chapter deals with analysis and interpretation of the data shown in Table 4.

Section 5.2 presents analysis of the data for experiments involving dichotomous noise bands. Feature detectabilities will be computed, and these values will be compared across experiments to evaluate the roles of interacting features. For example, if the response SNR for the dichotomous noise band in Experiment 3 is observed to be significantly greater than that in Experiment 1 given comparable values for P(C), results would suggest that the amplitude modulated fixed feature was acting as a confusion parameter. This finding would provide support for one of the hypotheses listed in Section 4.2.

Experiments involving amplitude modulation as the dichotomous feature constitute the topic of Section 5.3. Here, data will be used to compute the Weber fractions associated with the modulation, by noting the magnitude of intensity increments for the pure signal and the SNR where terminal decisions were made. These results will be

used to test the hypotheses concerning amplitude modulation, AM, as a dichotomous feature.

In Section 5.4, results from the previous two sections are compared with those predicted by the information processing model discussed earlier. Obvious differences between theory and experiment in some cases reflect problems with the experimental technique, while in others, they suggest revision of the theory. Numerous directions for further study are pointed out by these data, and these areas are addressed briefly in Section 5.5.

5.2 Analysis of Results for Noise Bands Dichotomous

Eleven of the fifteen experiments listed in Tables 3 and 4 involve dichotomous noise bands. The 1/3-octave spectra of four representative experiments are shown in Figures 5 to 8. These figures show the levels of signals, S_{H0} and S_{H1} above the background noise at the mean response SNR for the feature present case. In each figure, signal S_{H0} is shown at the same level as S_{H1} for comparison purposes only. In fact, subjects often responded at a different SNR when the probe stimulus did not contain the dichotomous band.

Figure 5 corresponds to the one-third-octave spectra for Experiment 1. This is the simplest discrimination task to be reported, with both signals containing an octave band of noise centered at 4 kHz and S_{H1} containing an octave band centered at 500 Hz. Figure 5 also shows the 1/3-octave spectra for Experiments 3 and 4 since the presence of amplitude modulation in these experiments does not affect

the spectral shape of the signals. Although the spectra for these latter experiments have the same shapes as those shown, the response SNR'S are not the same as that depicted in the figure.

Figure 6 illustrates the signal excess above the noise for Experiment 10. In this case, the fixed feature is a band of noise centered at 1 kHz, which is adjacent in frequency to the dichotomous 500-Hz band. Therefore, S_{H1} consists of a continuous band of frequencies which is two octaves wide.

The signals for Experiment 13 are illustrated in Figure 7. This experiment involves a dichotomous octave band centered at 1 kHz and two nonadjacent fixed features centered at 250 Hz and 4 kHz. Finally, Figure 8 shows spectral plots for Experiment 14 which involves two dichotomous noise bands and a single fixed noise band. This figure also represents the spectra for Experiment 16 in which the fixed 1-kHz band is amplitude modulated. However, the response SNR for Experiment 16 is different from that shown in Figure 8.

From the apperance of Figures 5 to 8, it seems reasonable to regard the dichotomous features as having bandwidths which are wider than one octave. At the response point, the signals are far enough out of the noise that some energy outside the octave bands may contribute to the feature detectabilities. In fact, values of d'_{mt} were calculated both by assuming effective rectangular bandwidths of one octave, and by assuming somewhat wider bandwidths. The wider bandwidth calculations resulted in larger values for d', but the

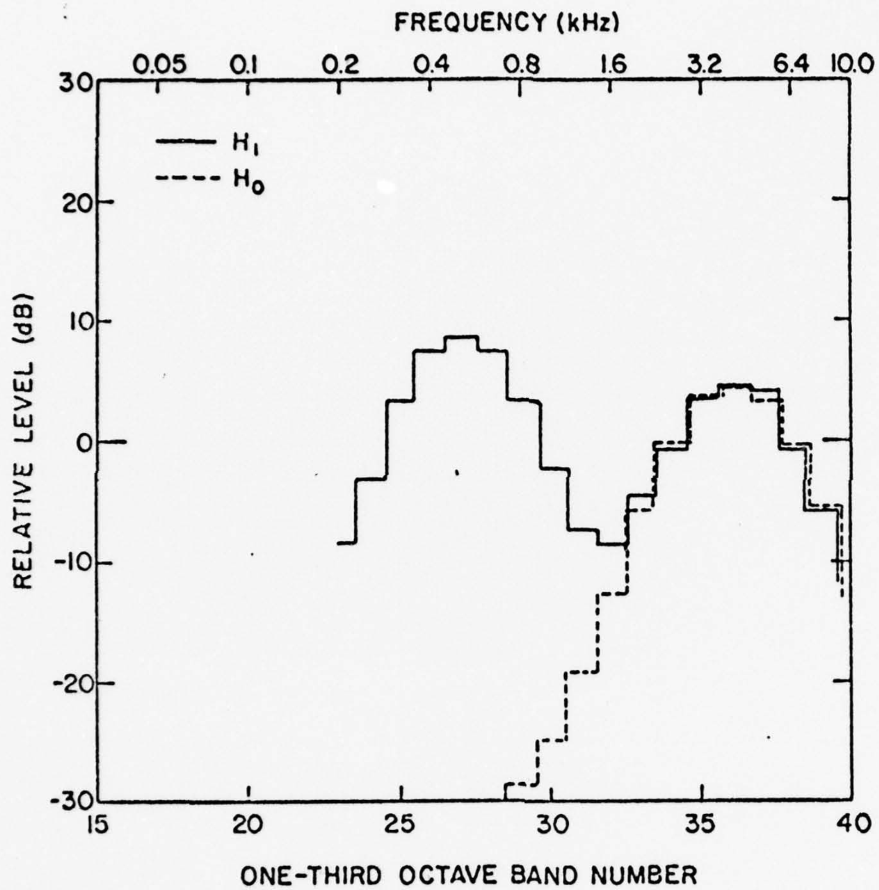


Figure 5. Signal Excess in One-Third Octave Bands at the Terminal Decision for Experiment 1, Dichotomous Noise Band with One Nonadjacent Fixed Noise Band.

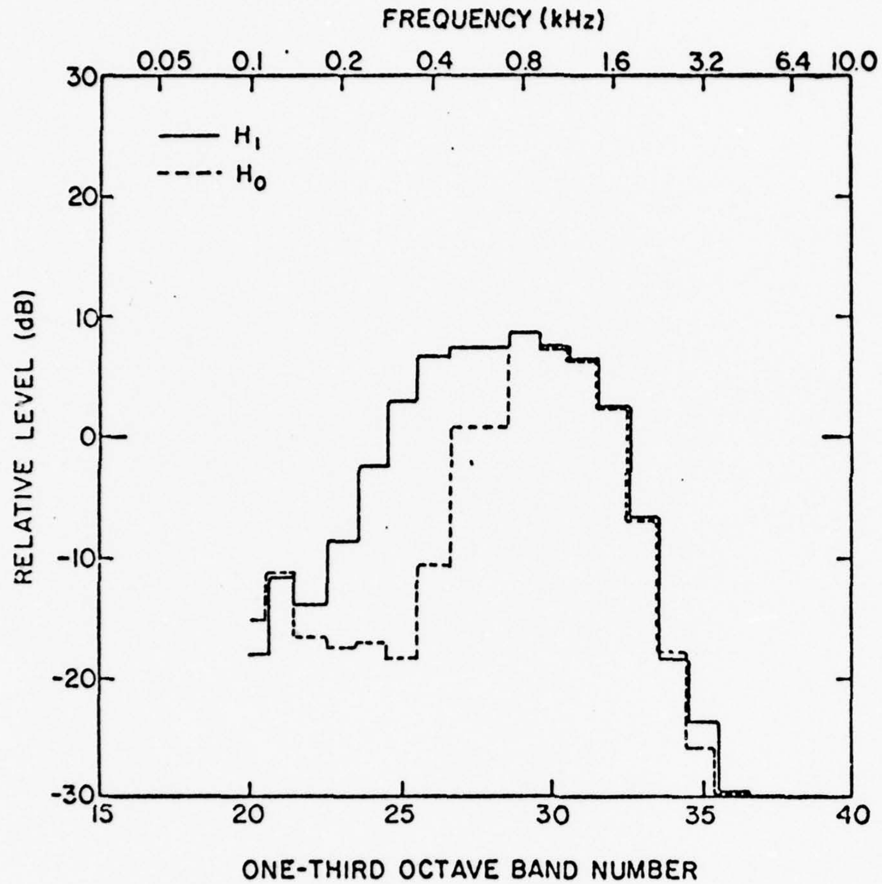


Figure 6. Signal Excess at the Terminal Decision for Experiment 10, Adjacent Fixed Noise Band.

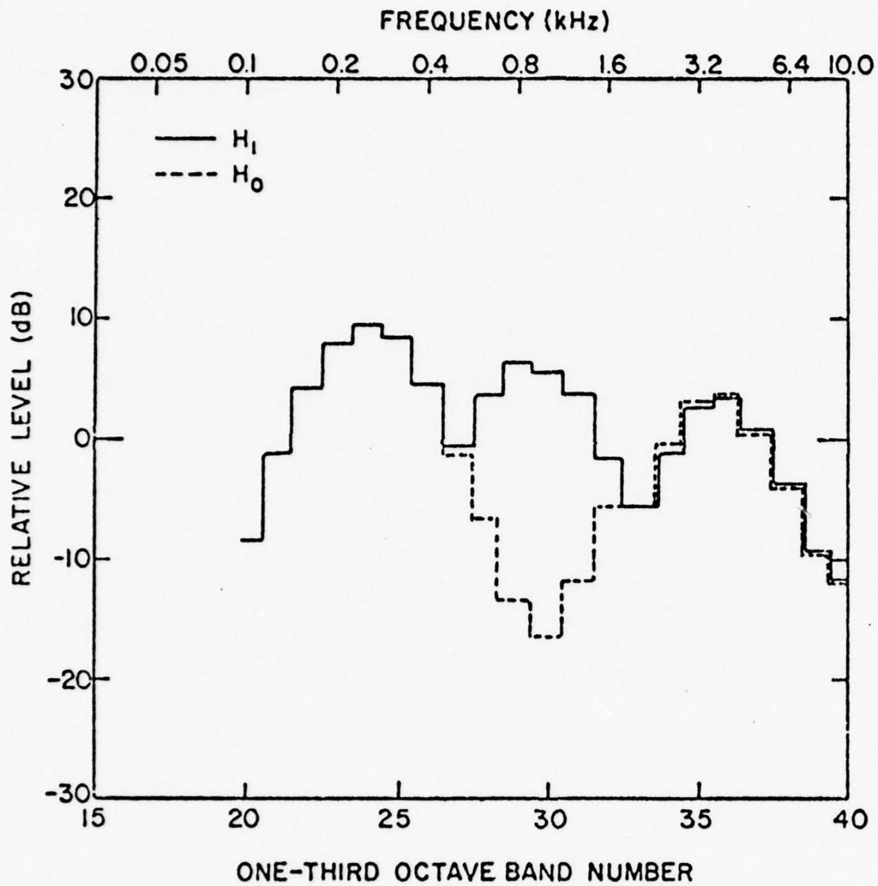


Figure 7. Signal Excess at the Terminal Decision for Experiment 13, Two Nonadjacent Fixed Noise Bands.

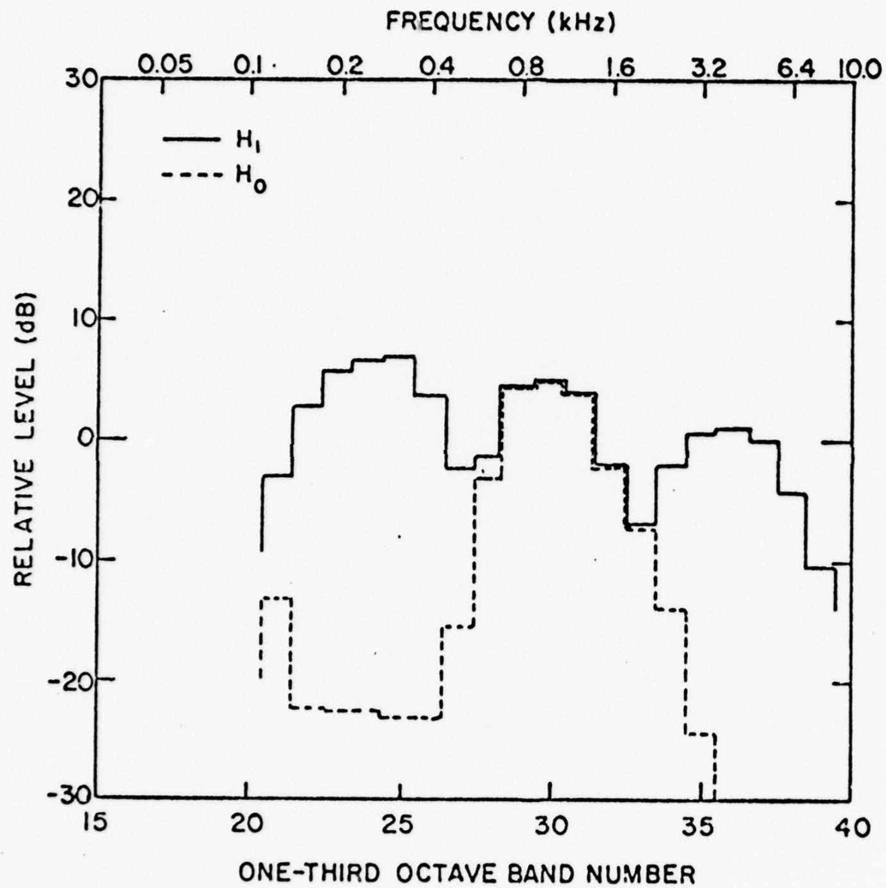


Figure 8. Signal Excess at the Terminal Decision for Experiment 14, Two Dichotomous Noise Bands.

relations among experiments remained nearly constant whichever method was used. The use of wider feature bandwidths did not affect the qualitative results to be reported, and therefore, the dichotomous features have been assumed to be one octave wide.

The results of detectability calculations for the eleven experiments to be discussed here are shown in Table 5. The 90% confidence limits on d'_{mt} are also shown in the table. These were determined using Equation (10) for the confidence limits on SNR. It has been assumed that there is no variance in the product W*T for each signal. An integration time of 500 msec has been used, and the feature bandwidths are those shown in Table 5. Table 6 shows the 90% confidence limits on the probability of correct responses for experiments with noise bands dichotomous. In most cases, the number of events is large enough to estimate P(C) within $\pm$10% of the correct value.

Looking at the results in Table 5, the most obvious trend in the data reflects an increase in d' with feature bandwidth. This can be readily seen by comparing d' values between Experiments 1 and 6. These experiments are conceptually similar, both involving a 4-kHz fixed noise band and a low frequency dichotomous band. The 500-Hz band in Experiment 1 has a significantly higher d' value than does the 250-Hz band in Experiment 6. Furthermore, the highest observed values of d' occur for Feature 2, a noise band centered at 4 kHz.

TABLE 5

SUMMARY OF DETECTABILITY INFORMATION FOR EXPERIMENTS INVOLVING DICHOTOMOUS BANDS OF NOISE

Experiment	Feature	W_{eff} (Hz)	10 log d'_{mt}	90% Confidence Interval d'_{mt}
1	1	354	11.93	11.77-12.07
3	1	354	12.22	12.12-12.30
4	1	354	10.72	10.49-10.93
5	1	354	11.75	11.57-11.90
6	5	177	9.81	9.49-10.08
8	5	177	10.74	10.04-11.28
9	2	2828	15.43	15.12-15.70
10	1	354	11.78	11.59-11.95
13	6	707	12.74	12.47-12.98
14	2	2828	13.81	13.33-14.26
14	5	177	10.14	9.93-10.31
16	2	2828	16.10	15.91-16.26
16	5	177	10.91	10.85-10.95

TABLE 6

SUMMARY OF PROBABILITIES EXPERIMENTALLY MEASURED FOR DISCRIMINATION WITH DICHOTOMOUS BANDS OF NOISE

Experiment	Correct Responses	P(C)	90% Confidence Interval P(C)
1	41/48	0.854	0.758-0.929
3	51/62	0.823	0.734-0.897
4	57/58	0.983	0.942-0.999
5	42/50	0.840	0.744-0.917
6	33/34	0.971	0.903-0.999
8	22/26	0.846	0.710-0.945
9	43/70	0.614	0.514-0.710
10	38/42	0.905	0.815-0.967
13	28/57	0.491	0.380-0.602
14	48/50	0.960	0.900-0.993
16	41/44	0.932	0.854-0.981

According to Green (1960a), the detectabilities of two signals having different bandwidths should be the same given equal values of P(C). That is, wide bandwidth signals should be detectable at lower SNR'S than narrow bandwidth signals, such that signal energies are equivalent. This was assumed to be true in designing the experiments reported here. The fact that the largest values of d' are observed for the widest bandwidth signals is quite disturbing since it seems to contradict accepted theory. Unfortunately, this will seriously limit the types of analysis which are possible with the present data. The detectabilities of dichotomous features having different bandwidths cannot be quantitatively compared, a fact which was not anticipated in the design of the experiments. No obvious empirical relation exists in the data for predicting observed detectability as a function of feature bandwidth. Analysis must therefore be limited to cases where two experiments involve dichotomous features having the same bandwidth. However, a number of meaningful results are still obtainable despite these restrictions.

The cause of the unexpected dependence of d' on feature bandwidth is probably related in some way to the experimental procedure. A partial explanation for this phenomenon may be provided by the following: In the experiments, the signals to be detected are the result of filtered white noise, and they are masked by a white noise background. In addition to acting as a masking stimulus, the background noise may also act as a confusion parameter to the extent that

it sounds like the features to be detected. This phenomenon, like that of signal detectability, would be a function of SNR. At higher SNR'S, subjects would be less likely to confuse the sounds of signals and background noise. Then, when narrow bandwidth features would become detectable based on energy considerations, this confusion phenomenon would probably not be an important factor due to the high response SNR. However, broadband features become detectable at lower SNR'S, based solely on energy considerations. At these levels, where the subject would normally respond in an experiment such as Green's, confusions between the sounds of the signals and noise may cause him to defer his decision. Then, the response SNR would reflect some combination of signal energy and confusion reduction.

During the experiments, one subject referred to as P. C. was asked to record the strategies he used in making decisions. In a large number of cases, he reported being confused by the signals "sounding like the background." He stated that the difference between signals in the exposure sets was obvious, but that perceptual confusions between the features and the background often interfered with his detection criterion. The fact that Green (1960a) did not observe this type of confusion phenomenon may be due to differences in experimental procedures. The use of a ramped SNR in a free-response setting, as well as the maintenance of stimuli at an equal loudness throughout the ramp, could contribute to differences between the present results and those observed by Green. This explanation

is, however, incomplete at best since it still fails to account for the difference between Experiments 1 and 6. The response to the 250-Hz band was actually at a lower SNR than the response to the 500-Hz band in Experiment 1.

Although the unexpected dependence of detectability on bandwidth causes problems in analysis, meaningful results can be obtained for features having the same bandwidth. Experiments 1, 5, and 6 are the simplest discrimination tasks to be reported, and they serve as a basis for comparison with the other experiments. Experiments 1 and 5 are identical, except that the dichotomous band in Experiment 5 has a lower spectral level relative to the fixed feature. As can be seen from Tables 5 and 6, performance on these two tasks is quite similar. The fact that changing the relative spectral levels of the features does not affect performance indicates that interactions between features have no great effect on discrimination performance. Janota (1977) found that for these signals, the detectability of the dichotomous noise band agreed well with results obtained by Green (1960a). The presence of the fixed noise band does not appear to affect detection of the dichotomous band. Subject P. C., in reporting his strategies, mentioned that he was listening for a low frequency band in these experiments. He stated that the experiments were quite easy, except for similarity confusions between the signals and the background.

Due to the bandwidth effect noted above, the observed d' in Experiment 6 is lower than the values in Experiments 1 and 5. However,

it is still reasonable to assume that feature interactions do not affect discrimination performance on this task. Therefore, Experiment 6 will be used as a basis of comparison for experiments having the 250-Hz noise band dichotomous. Likewise, results for Experiment 1 and 5 will be compared with those for experiments having the 500-Hz band dichotomous.

One important question which these experiments sought to answer is that concerning the interactive effect of fixed features which are adjacent in frequency to a dichotomous feature. To analyze this case, the observed detectabilities of features can be compared between Experiments 1 and 10. Both of these experiments involve a dichotomous octave band of noise centered at 500 Hz. In Experiment 1, the fixed noise band is centered at 4 kHz and is therefore widely separated in frequency from the dichotomous feature. In Experiment 10, the fixed band is centered at 1 kHz such that the fixed and dichotomous bands form a continuum of frequencies. The discrimination task in Experiment 10 therefore reduces to a problem in detecting a change in signal bandwidth under the two hypotheses. This can be seen graphically by comparing the spectra of Figures 5 and 6.

It was hypothesized in Section 4.2 that the fixed and dichotomous bands would be perceptually grouped, and that the adjacent feature would interfere with the detection of the dichotomous band. If this were true, the observed value of d' for the 500-Hz band should be higher in Experiment 10 than in Experiment 1. The observed values of

d' listed in Table 5 show clearly that this is not the case. That is, the adjacency of the fixed noise band does not result in a significant shift in d'. Table 6 shows that the probabilities of correct responses are comparable for the two experiments. In Experiment 10, the notes taken by Subject P. C. indicate that he was listening for a low frequency noise band, but that detection of the dichotomous feature required extreme concentration. No such difficulties were reported by the subject in Experiments 1, 5, and 6. However, his response SNR in Experiment 10 does not reflect this reported difficulty.

Another test of the adjacency hypothesis is provided by a comparison of Experiments 6 and 8. Both contain a 250-Hz dichotomous noise band and a 4-kHz fixed noise band. In addition, Experiment 8 contains an adjacent fixed noise band centered at 500 Hz. This case is certainly more complex than that depicted in Experiment 10, in that is involves multiple fixed features with one being adjacent to the 250-Hz band. Tables 5 and 6 show that the value of d' for the dichotomous feature is higher in Experiment 8 than in Experiment 6, while the P(C) is lower. However, these differences are not statistically significant at the 90% level.

The complicating effect of multiple fixed features will be discussed below. However, the comparisons between Experiments 1 and 10 and between Experiments 6 and 8 clearly show that an adjacent fixed feature does not interfere with discrimination in these tasks. The irrelevant feature information does not appear to hinder

discrimination performance. Further evidence to this effect will be given in Section 5.3, where it will be shown that the detection of amplitude modulation is not affected by the presence of an adjacent fixed noise band.

The problem of detecting a noise band in the presence of multiple fixed features is handled in Experiments 8 and 13. In Experiment 13, the dichotomous feature is an octave noise band centered at 1 kHz, and the signals contain two nonadjacent fixed features centered at 250 Hz and 4 kHz. In Experiment 8, the dichotomous feature is a 250-Hz band, and one of the fixed features is adjacent. Unfortunately, comparisons between these and other experiments are difficult since the feature bandwidth problem prevents quantitative analysis. Some qualitative results concerning multiple fixed features are, however, worthy of discussion.

Referring to Table 6, it can be seen that the probability correct for Experiment 13 is significantly lower than that for all other experiments. This experiment, which involves a dichotomous band located between two nonadjacent fixed bands, appears to have been very difficult. Since subjects only responded at chance level, it appears that the fixed features interfered with discrimination. Unfortunately, values of d' cannot be quantitatively compared. Subject P. C. reported that this experiment was indeed very difficult, and he apparently used completely different strategies each time he performed the task. On one occasion, he reported "listening for a hollow

sound, with something missing in the mid-frequencies." On another occasion, he reported listening for low frequency noise bands, while in a third case, he was listening for high frequency bands. The inability to establish a consistent discrimination strategy, in addition to the responses at chance level for all subjects suggest that the detection of a dichotomous feature located between two fixed features is extremely difficult. The interactive effect of irrelevant information both above and below the dichotomous band seems to seriously confound the discrimination task.

As reported above, Experiment 8 involved multiple fixed features, but both were at higher frequencies than the dichotomous feature. When results were compared between Experiments 6 and 8, no statistically significant differences were found. Subject P. C. stated that this experiment was not difficult, except that the dichotomous feature sounded like the background noise.

One additional experiment was conducted which concerned possible interactive effects from a fixed band of noise. Experiment 9 involved a dichotomous band centered at 4 kHz and a low frequency fixed band which spanned two octaves. It was hoped that this experiment would provide information about the possible interactive effect of changing the bandwidth of an irrelevant feature. However, due to the dependence of detectability on bandwidth, a basis of comparison for Experiment 9 does not exist in the data. Furthermore, analysis of the strategy employed by Subject P. C. does not provide insight leading to a

qualitative interpretation of the data. Therefore, no conclusive statements can be made from the data regarding the effect of the fixed feature bandwidth.

Finally, the discussion of fixed feature interactions with a dichotomous band of noise concludes with an analysis of amplitude modulation as an irrelevant feature. It was hypothesized in Section 4.2 that an obvious modulation component which is not relevant to the discrimination task would distract subjects' attention. That is, they would have some difficulty attending to a dichotomous band of noise in the presence of amplitude modulation. Experiments 3 and 16 were constructed to test this hypothesis. The results are shown in Tables 5 and 6.

Experiment 3 involved a dichotomous noise band centered at 500 Hz and a fixed band centered at 4 kHz which was amplitude modulated. Thus, the difference between Experiments 1 and 3 consists solely of the amplitude modulation. It can be seen from the data that the value of d' for the noise band is significantly higher in Experiment 3 than in Experiment 1. Meanwhile, the observed probabilities of correct responses are nearly the same in the two experiments. From these data, it is reasonable to conclude that the modulation interfered with subjects' ability to focus on the dichotomous band. The higher response SNR in Experiment 3 indicates that subjects were distracted by the irrelevant information.

Further evidence supporting this idea is provided by comparing results for Experiments 14 and 16. These experiments involved two

dichotomous noise bands, one above and the other below the frequencies in a fixed band centered at 1 kHz. In Experiment 16, this fixed band was modulated by a 10-Hz square wave, while in Experiment 14, it was not. It can be seen from the results in Table 5 that both of the dichotomous features had higher d' values in Experiment 16 than did the corresponding bands in Experiment 14. Again, values of P(C) are similar for the two cases. A discussion of multiple dichotomous features will be presented below. However, whatever criterion subjects use to combine the information in these features, it is reasonable to assume that they use the same feature combination in both experiments. With this assumption, the data indicate quite clearly that amplitude modulation as a fixed feature interferes with discrimination performance. This is true whether the task involves a single dichotomous feature or multiple dichotomous features.

In describing his strategy for Experiment 16, Subject P. C. stated that the modulation was a nuisance. He stated that even though he knew from the exposure set that this feature was irrelevant, he initially paid attention to the modulation and ignored the noise bands. On one occasion, he described the dichotomous feature as consisting of low frequency noise, but on later trials with the same experiment, he stated simply that S_{H1} was "noisier" without reference to frequency.

Experiments 14 and 16 involve two dichotomous noise bands. Experiment 14 was constructed to answer questions about how subjects

combine featural information. Do they, as Fidell observed, rely solely on the most detectable feature? Or, as Green observed with tonal signals, is some combination of feature detectabilities involved (Fidell et al., 1974; Green, 1958)? Regarding the data in Table 5, one is tempted to conclude that decisions were based solely on the 4-kHz band, Feature 2, since this feature has the highest value of d' at the response point. However, this higher value of d' more likely reflects the bandwidth trend rather than providing any information about the discrimination task. Table 6 shows that the probability correct was very high, 0.960 for fifty events, so the task may be regarded as very easy.

Subject P. C. indicated on some occasions that he was listening for low frequency noise in order to perform the tasks in Experiments 14 and 16. However, on most trials he described the difference between the two signals in terms of a unified percept. That is, he found S_{H1} to be "rougher" or "to contain more noise" than S_{H0}. None of his comments reflected the idea of two separate noise bands. Unfortunately, no further inferences can be drawn from the data about how two dichotomous noise bands are perceptually combined in performing the discrimination tasks.

One additional experiment was conducted involving two dichotomous features. In Experiment 4, however, the dichotomous portion of the signal was a noise band centered at 500 Hz as well as amplitude modulation of that band. The experiment sought to determine whether

the noise band, the modulation, or some combination of features was most important in reaching a discrimination decision. A comparison of results between Experiments 4 and 1 reveals that the detectability of the noise band is significantly lower in Experiment 4. That is, subjects were willing to respond at a much lower SNR in Experiment 4 than in any other experiment involving a dichotomous band at 500 Hz. From this fact alone, it is likely that the noise band was not the major contributor to the discrimination decision. Furthermore, as will be shown in the next section, the observed value of the Weber fraction in Experiment 4 agreed favorably with values for other experiments involving amplitude modulation dichotomous. Finally, Subject P. C. described the difference between sounds solely in terms of the modulation, with no mention of the dichotomous noise band. From this combination of facts, it can be inferred that the discrimination decision was based on the amplitude modulation and not on the noise band.

5.3 Analysis of Results for Amplitude Modulation Dichotomous

Five experiments were conducted in which the dichotomous feature was amplitude modulation of a noise band centered at 500 Hz. In these experiments, Experiments 2, 4, 11, 12, and 15, discrimination involved detection of the modulation in the presence of five different fixed features. Under the assumption that this modulation is perceived as separate intensity increments, the detection criterion is based on the

ratio $\Delta I/I$, the Weber fraction (Janota, 1977; Moore and Raab, 1975). As the signal-to-noise ratio increased during an experimental trial, the Weber fraction also increased until conditions were such that subjects were able to respond. Then, the possible interactive effects of invariant features will be interpreted in terms of the observed intensity ratios in the five experiments.

In all five experiments, a band of noise centered at 500 Hz was modulated by a 10-Hz square wave having a 50% duty cycle. Thus, the modulation was periodic, with the duration of the intensity increments being 50 msec. The bandwidths associated with the dichotomous modulation were designed to be one octave. However, since the roll-off of filters used to create the signals was not infinitely steep, relevant information existed outside this octave in some experiments. The second column of Table 7 lists the appropriate feature bandwidths for the experiments.

In order to calculate the observed Weber fraction for each experiment, two measured quantities are needed. These are the mean SNR to respond, listed in Table 4, and the ratio of peak to average intensity for the pure signal, listed in the third column of Table 7. The difference between peak and average intensities for the signals without interfering noise was measured using the equipment described in Section 4.5. Measurement errors associated with this quantity may be as large as 1/2 dB.

Weber fractions for the five experiments were calculated using a method suggested by Janota (1977). Results for the mean are shown in Table 7 along with the 90% confidence limits on this quantity. These confidence limits only reflect the variance in response SNR, and they do not include errors associated with the measurement of peak and average intensities. A sample Weber fraction calculation using the data for Experiment 11 is given below. This experiment simply involved a noise band which was modulated in one case and not in the other. No additional fixed features were involved.

Referring to Table 4, the response SNR for Experiment 11 in the 500 Hz band was -0.84 dB. The difference between the peak and average spectral levels for the signal without interfering noise is given in Table 7 as 3 dB. Using these quantities, the following can be calculated:

1. At the response point, the ratio of the average signal-plus-noise to noise is given by:

   -0.84 dB + 0 dB = 2.61 dB.

Note that dB's are combined such that the addition of two equal levels results in an increase of 3 dB.

2. The ratio of signal-plus-noise to noise at the peak of the envelope amplitude excursion is:

   2.16 dB + 0 dB = 4.22 dB.

3. The ratio of S+N/N at the peak to the average S+N/N is then 1.61 dB.

TABLE 7

EXPERIMENTAL RESULTS FOR DISCRIMINATION TASKS INVOLVING AMPLITUDE MODULATION AS A DICHOTOMOUS FEATURE

Experiment	W_{eff} (Hz)	Signal I_p/I_{av} (dB)	$\Delta I/I$ (dB)	90% Confidence Interval ($\Delta I/I$)
2	354	2	-4.34	-4.44 to -4.20
4	354	2	-3.84	-4.02 to -3.71
11	536	3	-3.47	-3.85 to -3.13
12	354	3	-3.29	-3.68 to -2.98
15	536	3	-0.18	-0.24 to -0.14

4. Therefore, the ratio $I+\Delta I/I = (1.61/10) = 1.45$.

5. Therefore, $\Delta I/I = 0.45$.

6. Then, $10 \log (\Delta I/I) = -3.47$ dB. This quantity is listed as the mean value of the Weber fraction in Table 7.

Table 8 gives the observed probability of correct responses for the five experiments, along with the 90% confidence limits on P(C). It can be seen that values of P(C) for the first four experiments approach 100%, indicating that subjects had little trouble with these discrimination tasks. Subject P. C. reported that these four experiments were extremely easy, i.e., that the modulation was very obvious. When the dichotomous feature was absent, he stated that his decision was based on the elapsed time since the beginning of the event. That is, he usually did not rely on detection of the fixed features in order to respond in the feature absent cases. Rather, his decisions in these cases were of the form, "If the modulation were present, I should have heard it by this time."

The results in Table 7 reveal that the observed Weber fractions for Experiments 2, 4, 11, and 12 are not very different from one another. There is a statistically significant difference between the performance for Experiment 2 and that for the other experiments, but this difference is probably not meaningful due to the large measurement error in the ratio of peak-to-average intensities for the pure signals. This difference, if significant, would indicate that the 4 kHz noise band in Experiment 2 caused a slight improvement in performance as

TABLE 8

SUMMARY OF PROBABILITIES EXPERIMENTALLY MEASURED FOR DISCRIMINATION WITH AMPLITUDE MODULATION DICHOTOMOUS

Experiment	Correct Responses	P(C)	90% Confidence Interval P(C)
2	108/108	1.00	0.993-1.0
4	57/58	0.983	0.942-0.999
11	55/55	1.00	0.987-1.0
12	22/22	1.00	0.968-1.0
15	48/53	0.906	0.828-0.962

compared to the other experiments. However, this interpretation of the data is doubtful, and it will therefore be assumed that differences among the first four experiments in Table 7 are not significant.

It can be seen that the value of $\Delta I/I$ measured in Experiment 4 is in good agreement with values for Experiments 2, 11, and 12. This result further supports the contention in Section 5.2 that it was the modulation and not the dichotomous noise band which contributed most to the discrimination decision. The modulation was very perceptible to the subjects, and they appear to have used this feature almost exclusively in reaching their decisions. The result for Experiment 4 is difficult to generalize, however, because a less obvious modulation component may force subjects to use the information in the dichotomous noise band.

Experiments 2, 11, and 12 were designed to test the interactive effect of a fixed noise band on the detection of modulation. Experiment 11 consisted solely of a 500-Hz band which was modulated in one case and not in the other. Experiment 2, in addition to the dichotomous modulation, contained a nonadjacent fixed noise band centered at 4 kHz. Experiment 12 involved an adjacent fixed noise band centered at 1 kHz. As indicated above, the results in Table 7 show that performance on these three tasks is nearly equivalent. The presence of a nonadjacent fixed noise band does not appear to affect a subject's ability to detect the dichotomous modulation. A comparison between Experiments 11 and 12 also shows that the presence of an adjacent

noise band has little effect on performance. If any interactions exist between the fixed and dichotomous features in these three tasks, their effects seem to be negligible. Again, the observed probability of correct responses indicates that subjects had little trouble with these discriminations.

Referring again to Table 7, the results for Experiment 15 differ significantly from those for the other experiments. This task involved the detection of modulation in the presence of modulation. The dichotomous feature was, as before, modulation of the 500-Hz band, but in this experiment, the fixed feature consisted of a modulated 4-kHz band. Both noise bands were modulated by the same square wave such that the ratio of peak to average intensities was the same in both bands. Subjects were asked, in effect, to discriminate between modulated signals of different bandwidths.

Since the ratio of peak-to-average intensities for the pure signal was 3 dB, the maximum attainable value for the Weber fraction would be 0 dB. That is, the ratio $\Delta I/I$ at infinite signal-to-noise ratio is 0 dB. In Experiment 15, the SNR at the response point was 14.13 dB, resulting in an observed Weber fraction of -0.18 dB. This high response SNR shows that subjects were receiving almost no new information in the last few dB of increase prior to the response point. They could therefore have responded several dB earlier with no loss of information. This fact is reflected in the large standard deviation in response SNR of 5.41 dB shown in Table 4. Due to the high response

SNR, this large variance reduces to a quite small confidence interval on the Weber fraction.

It is not clear why subjects did not respond earlier on these trials since they gained very little information by deferring their decisions. Their performance was very accurate, however, with an observed P(C) of 0.91 as shown in Table 8. Despite the high percentage of correct responses, it can be inferred from the Weber fraction that Experiment 15 was a very difficult task. Amplitude modulation as a fixed feature seems to interfere with the discrimination of AM dichotomous, just as it did with the discrimination of dichotomous noise bands. The clear difference between performance on Experiment 15 and that on the other experiments in Table 7 shows that subjects had difficulty attending to the relevant feature in this task. Had they been able to apply a narrowband filter to the signals, as would have been the case for an ideal observer, performance would not have differed between Experiment 15 and the others. Subject P. C. did describe this task as very difficult, although he appeared to be using the correct strategy, listening for modulation of the low frequency band.

5.4 Comparison Between Results and the Model

Following a review of the literature on auditory discrimination tasks, signal detection theory, and auditory information processing, a theoretical model for the discrimination of noise-like sounds was developed in Section 3.4. This model includes feature extraction and

matching of features to remembered patterns as integral parts of the information processing system. The model differs from others in the literature in that it attempts to account for interactions between features which might aid or hinder discrimination performance.

A list of hypotheses about how subjects discriminate between noise-like sounds was developed from the model. Then, sixteen experiments were designed to test these hypotheses and consequently test various aspects of the model. Results of these experiments constitute the topics of Sections 5.2 and 5.3. These results were only analyzed for cases where the probe stimulus contained the dichotomous features. Major conclusions drawn from the data and from commants by a subject about the strategies he used in performing the tasks are the following:

1. A nonadjacent, fixed noise band does not affect the discrimination of a dichotomous feature when the dichotomous feature is either a band of noise or amplitude modulation. That is, no significant interactions were observed between fixed and dichotomous features. In fact, previous research has shown that performance on these types of discrimination tasks may be accurately modeled in terms of detecting a dichotomous feature, and furthermore, feature detectabilities are in good agreement with those for tasks involving no fixed features (Janota, 1977).

2. An adjacent fixed noise band does not interfere with discrimination when the dichotomous feature is either a noise band or amplitude modulation. No significant differences in discrimination

performance were found when comparing cases with adjacent and non-adjacent fixed features.

3. Amplitude modulation as an irrelevant feature was found to degrade performance for cases of a dichotomous noise band, two dichotomous noise bands, as well as amplitude modulation as a dichotomous feature.

4. When discrimination involves detection of a dichotomous noise band in the presence of two fixed noise bands, one above and one below the frequencies of the dichotomous feature, the fixed features degrade discrimination performance.

5. When signals involve two dichotomous noise bands, the perceived difference between them is a unified percept rather than two separate bands. However, the extent to which each band contributes to the perceived difference cannot be determined from the present data.

6. When signals involved both amplitude modulation and a noise band as dichotomous features, the discrimination decisions were based on perception of the modulation. The dichotomous noise band contributed very little to the decisions.

7. In addition to acting as a masking stimulus, the background noise in many cases also acted as a confusion parameter to the extent that it sounded like the features to be detected. The extent to which this confusion phenomenon affected discrimination performance is difficult to determine.

For cases where feature interactions do not strongly influence discrimination performance, Janota (1977) has shown that the detectabilities of dichotomous noise bands and the Weber fractions associated with dichotomous modulation, agree well with the results of classical detection experiments. This provides strong evidence for the feature extraction stage of the discrimination model followed by hypothesis tests on the presence or absence of acoustic features. Beginning with the assumption that simple discrimination tasks could be analyzed in this manner, the model in Section 3.4 hypothesizes various types of feature interactions which either degrade or enhance discrimination performance on more complex tasks. The conclusions listed above show that some of these proposed interactions were observed in the data, while others were not.

The model proposes two ways in which an irrelevant acoustic feature might degrade discrimination performance. Since filter characteristics in the auditory system are not infinitely steep, it was hypothesized that some energy from an adjacent feature could contribute to the detectability computations for a dichotomous noise band. This type of interaction was not observed in the data. No significant differences were found between cases involving adjacent and nonadjacent fixed features. The possibility that subjects might set their filter cutoff frequencies somewhat below the upper edge of the dichotomous band does not seem adequate to describe their performance. However, it can be stated that they were somehow able to separate the two bands such that no important interactions occurred.

Another type of information leakage is hypothesized in the model, wherein the presence of a dominant but irrelevant feature interferes with a subject's ability to attend to the relevant signal information. This type of distraction was indeed observed in all cases where amplitude modulation was used as a fixed feature. These cases covered a range of signals involving three different types of dichotomous information. In all cases, the subjects needed a higher feature detectability in order to make the discriminations. Comments by Subject P. C. support this interpretation of the data. He stated frequently that he was attending to the modulation, even though he knew that it was not relevant to the task.

Two types of feature interactions which tend to degrade performance were observed in the data but not specifically shown in the model. First, when signals contained fixed features both above and below the frequencies of a dichotomous band, subjects seemed unable to accurately adjust the band-pass filters necessary to perform the task. Secondly, performance on a number of the tasks was degraded by confusions between similar sounding signals and noise. These interactions were proposed in the list of hypotheses in Section 4.2, and a more precise model of auditory discrimination should include them either in the feature extraction stage or in the hypothesis testing stage.

For the case of multiple dichotomous features, the model predicts that the hypothesis tests are not independent, and that the discrimination decision may be made when either or both features exceed some criterion.

This hypothesis could not be quantitatively analyzed with the existing data, but subjects' comments indicate that multiple dichotomous noise bands were perceived as a unit rather than as separate bands. The perfect correlation between these features did affect performance, but the magnitude of this effect cannot be determined with these data. The model also shows that performance should be facilitated by the presence of two dichotomous features. Although the percentage of correct responses was very high for this experiment, subjects appeared to use a very strict criterion so that most of the experiments gave high values for P(C).

Finally, performance on the experiment involving both a noise band and amplitude modulation as dichotomous features was dominated by the amplitude modulation. Since performance on this task did not differ significantly from that on tasks involving only amplitude modulation dichotomous, it can be concluded that subjects were ignoring the noise band. They were therefore not using all available relevant information, probably because the modulation was much easier to extract than was the noise band. It is possible that subjects' decision criteria were exceeded solely by the modulation. However, it is also possible that the learned patterns stored in memory were dominated by the modulation, such that subjects regarded the task as involving only modulation as a relevant feature.

Further experiments are needed to analyze the decision stage of the model, wherein it is hypothesized that the results of feature tests are compared with remembered patterns. A number of possible

problems exist in this area concerning the accuracy of the remembered patterns as well as the stability of subjects' criteria over time. The accuracy of the internally stored representations must be called into question, first, because the sounds were initially unfamiliar to the subjects. That is, they do not fall within the classes of stimuli to which subjects are regularly exposed. In addition, the differences between similar noise-like stimuli cannot be easily verbalized, a fact which is likely to influence a subject's memory.

As mentioned earlier, the modified threshold procedure does not allow the separation of time-related and detectability-related factors which influence criterion. It is difficult to determine the extent to which discrimination decisions were based on response time rather than feature detection, but the high probability of correct responses in the feature present cases suggests that subjects were indeed detecting the dichotomous features. However, in the feature absent cases, Subject P. C. often commented that his responses were based on the elapsed time from the beginning of the trial. In these cases, he stated that S_{HO} was the probe because in the elapsed time, he had not heard the relevant feature. Likewise, he had not detected the fixed features prior to his responses.

The high percentage of correct responses for most experiments indicates that subjects were generally applying extremely strict decision criteria. In earlier experiments using the same instructions, Janota (1977) observed that sonar operators and some student subjects

were willing to respond using much more lax decision criteria. The reasons for these differences are not obvious and warrant further investigation.

Possible biases which warrant further study may also affect the response stage of the model. Comments made by subjects, as well as the high percentage of correct responses suggest that subjects had little trouble associating the letters A and B with the signals. However, one response bias which was observed in subjects' comments concerns their interpretation of the instruction that signals would occur with equal probability. The listeners commented that they often gave a B response following two A responses because they doubted that the same probe would be used on three consecutive trials. The tapes were indeed constructed such that the same probe never occurred three times in a row. However, an incorrect response on a previous event could result in errors on later events if subjects fallaciously applied the probability rules in the manner suggested.

5.5 Areas for Further Study

The results presented in this thesis provide a wealth of new information about how human observers discriminate between complex noise-like sounds. The thesis has investigated ways in which feature interactions affect discrimination performance on tasks involving acoustic features which are present in one signal and absent in the other. Some of the hypothesized interactions were observed in the data, while others were not. In addition to answering a number of

relevant questions in this area, the work has suggested a number of topics which require further study in the analysis of discrimination performance. Some of these topics relate to the feature extraction and hypothesis testing aspects of the proposed discrimination model. These areas reflect possible feature interactions which could not be analyzed with the present data. Some additional topics requiring further investigation relate to procedural considerations. These include studies of subjects' decision criteria and response biases which may influence results obtained with the modified threshold technique.

The data collected for this thesis was inadequate to analyze a number of important problems concerning feature interactions. Some of the results were of marginal utility in that they did not directly address the hypotheses which they were designed to test. In addition, some of the analysis pointed out new areas for investigation which were not originally considered to be important. Among the problems needing further study are the following:

1. The dependence of detectability on feature bandwidth using the modified threshold procedure. As discussed in Section 5.2, wide bandwidth features had higher values of d' at the response point than did narrow bandwidth features, a fact which seems contradictory to the idea that equidetectable features should have equal energies. A number of comparisons between experiments were prevented by this unanticipated problem. Determination of the magnitude and causes of this effect would aid the analysis of this type of discrimination task.

2. The extent to which background noise acts as a confusion parameter in the discrimination of noise-like stimuli. Comments by subjects supported the idea that confusions resulted because the background noise sounded like some of the broadband features. It was suggested that this confusion phenomenon is a function of SNR, and a method of measuring the extent to which it influences discrimination performance should be developed.

3. The ways in which feature detectabilities combine in the analysis of tasks involving multiple dichotomous features. In several tasks with two dichotomous noise bands, subjects perceived the difference between stimuli as a unit. However, it is not known how much each feature contributed to this unified percept.

4. The effect of fixed feature bandwidth on discrimination performance. An experiment was designed to answer questions in this area, but the necessary comparisons were prevented by the undetermined relation between detectability and bandwidth.

Finally, in order to properly interpret discrimination results obtained with the modified threshold procedure, a greater understanding is needed of the factors influencing a subject's decision criterion. The following is a list of topics which require further experimentation in this area.

1. Separation of time-related and detectability-related criterion effects. An experimental method is needed which allows independent analysis of these variables. With the modified threshold procedure,

time-related variables such as memory decay cannot be separated from variables which affect a subject's detection criterion.

2. The accuracy of remembered patterns. Due to subjects' unfamiliarity with the sounds and their inability to describe them verbally, it is not known how the accuracy of remembered patterns differs from that in studies with more familiar stimuli. The extent to which memory accuracy affects a subject's decision criterion is also an important factor in the understanding of discrimination performance.

3. Differences between strict and lax criterion used by subjects given the same instructions. It is not known why subjects in the present experiments required a much higher degree of certainty before responding than did earlier subjects.

4. Effects of response biases. Various types of response biases have been suggested in this thesis and in studies by other investigators (Janota, 1977; Cornell, 1978). Evaluation of these biases certainly warrants further consideration in the design of experiments and the interpretation of results with human subjects.

CHAPTER VI

SUMMARY AND CONCLUSIONS

The objective of the experiments reported in this thesis was to analyze the effects of feature interactions on discrimination performance with complex noise-like sounds. Noise-like sounds were defined as any sounds other than speech or music which could potentially convey information to a listener. Discrimination between such sounds plays an important role in a large number of industrial settings where an observer might rely on the sound of a machine to judge its operating characteristics or to detect possible malfunctions.

Sixteen pairs of laboratory-generated sounds were used for the experiments. Within a given sound pair, signals differed by one or more "dichotomous features," features present in one signal and absent in the other. The acoustic features used to compose the signals included octave bands of noise centered at various frequencies as well as amplitude modulation of noise bands by a 10-Hz square wave. Discrimination performance was studied under various conditions of "fixed" or irrelevant features, as well as several conditions involving multiple dichotomous features. The experiments were designed to show the extent to which the detectability of a given dichotomous feature was affected by interactions within the acoustic environment.

Chapter II of the thesis presents a brief review of relevant literature on auditory discrimination tasks, including tonal stimuli, speech recognition, noise-like sounds, as well as amplitude modulation. Chapter III presents the theoretical basis for the present studies, which involved pertinent aspects of the theory of signal detectability and information theory as applied to auditory processing. A theoretical model for predicting discrimination performance was developed using the experimental results presented in Chapter II and the theoretical considerations discussed in Chapter III.

The basic assumptions of the model were:

1. That detection of a dichotomous feature was a necessary condition for discrimination on the tasks of interest.

2. That the detectability of a given dichotomous feature would depend on interactions between features.

The model consisted essentially of a feature extraction stage followed by hypothesis tests about the presence or absence of relevant features. Interactions which either enhance or degrade discrimination performance were shown as affecting the computations of feature detectabilities. A list of hypotheses was derived from the model which can be summarized as follows:

1. When basing a discrimination decision on the detection of a dichotomous feature, the discriminability of that feature will be largely determined by the acoustic environment which includes such factors as the number and bandwidths of fixed features as well as the

extent to which the masking noise sounds like the dichotomous feature.

2. When signals differ by two or more dichotomous features, the way in which the component detectabilities combine will be determined by the nature of the discrimination task. However, the task will be easier than any of the cases where signals differ by only one of these features.

3. If a dichotomous band of noise is adjacent in frequency to a fixed feature, the two will be perceptually grouped, and the fixed feature will tend to degrade discrimination performance.

4. When signals involve several dichotomous features, one of which is amplitude modulation, the perceptual difference between sounds will be dominated by the modulation, and the discrimination decision will therefore be based on detection of this feature.

5. When two signals to be discriminated involve amplitude modulation as an irrelevant feature, it will distract the subjects and therefore degrade performance.

6. In general, the data will lend further support to a feature extraction model of complex sound discrimination.

In Chapter IV the experiments designed to test the above hypotheses were discussed. Five graduate students served as subjects for the studies which lasted approximately seven months. The tests involved sixteen sound pairs and were conducted using a procedure called the "modified threshold technique." In the procedure, subjects were presented with two signals designated "A" and "B," and one of the two then appeared as the probe in a white noise background.

The signal-to-noise ratio was increased slowly with time until subjects were willing to respond under the condition that they were "reasonably certain" of a decision. The important measured quantities were the SNR to respond and the probability of a correct response.

After the deletion of some unreliable data such as training tapes, results for the feature present probe condition were pooled across subjects. Results for dichotomous noise bands were normalized to a common base by computing the detectability in each experiment. Values of the Weber fraction were computed for experiments involving dichotomous modulation. Comparisons between experiments were limited to some extent by an unexpected dependence of detectability on feature bandwidth. Nevertheless, analysis of the data led to the following conclusions:

1. A fixed noise band which is far-removed in frequency from a dichotomous feature does not affect the discrimination of a dichotomous feature when the dichotomous feature is either a band of noise or amplitude modulation. That is, no significant interactions were observed between fixed and dichotomous features.

2. A fixed noise band which is adjacent in frequency to a dichotomous feature does not interfere with discrimination when the dichotomous feature is either a noise band or amplitude modulation. No significant differences in discrimination performance were found when comparing cases with adjacent and nonadjacent fixed features.

3. Amplitude modulation as an irrelevant feature was found to degrade performance for cases of a dichotomous noise band, two dichotomous noise bands, as well as amplitude modulation as a dichotomous feature.

4. When discrimination involves detection of a dichotomous noise band in the presence of two fixed noise bands, one above and one below the frequencies of the dichotomous feature, the fixed features degrade discrimination performance.

5. When signals involve two dichotomous noise bands, the perceived difference between them is a unified percept rather than two separate bands. However, the extent to which each band contributes to the perceived difference cannot be determined from the present data.

6. When signals involved both amplitude modulation and a noise band as dichotomous features, the discrimination decisions were based on perception of the modulation. The dichotomous noise band contributed very little to the decisions.

7. In addition to acting as a masking stimulus, the background noise in many cases also acted as a confusion parameter to the extent that it sounded like the features to be detected. The extent to which this confusion phenomenon affected discrimination performance is difficult to determine.

Results were compared with those predicted by the theoretical discrimination model, and it was seen that some of the hypothesized interactions occurred, while others did not. Several areas which warrant further study were suggested, including further experimentation with various types of multiple feature interactions as well as examination of ways in which the experimental procedure affects subjects' decision criteria.

REFERENCES

Békésy, G. von, <u>Experiments in Hearing</u>, New York: McGraw-Hill, (1960).

Corcoran, D. W. J., A. Carpenter, J. C. Webster, and M. M. Woodhead, "Comparison of Training Techniques for Complex Sound Identification," J. Acoust. Soc. Amer., 44:1, 1968, 157-167.

Corcoran, D. W. J., J. L. Dennett, and A. Carpenter, "Cooperation of Listener and Computer in a Recognition Task: Effect of Listener Reliability," J. Acoust. Soc. Amer., 47:2, 1970, 480-488.

Corliss, E. L. R., "Estimate of the Inherent Channel Capacity of the Ear," J. Acoust. Soc. Amer., 50:2, 1971, 671-677.

Cornell, P. T., "Detection of an Octave Band of Noise as a Function of Stimulus Presentation," The Pennsylvania State University, Masters Thesis, March 1978.

Creelman, C. D., "Detection of Signals of Uncertain Frequency," J. Acoust. Soc. Amer., 32, 1960, 805-810.

Dubrevski, N. A., and I. N. Tumarkina, "Investigation of the Human Perception of Amplitude-Modulated Noise," Sov. Phys. Acoust., 13:1, 1967, 41-47.

a. Ehmer, R. H., "Masking Patterns of Tones," J. Acoust. Soc. Amer., 31, 1959, 1115-1120.

b. Ehmer, R. H., "Masking by Tones vs. Noise Bands," J. Acoust. Soc. Amer., 31, 1959, 1253-1256.

Eriksen, C. W., and H. W. Hake, "Multidimensional Stimulus Differences and Accuracy of Discrimination," J. Exp. Psychol., 50, 1955, 153-160.

Fidell, S., K. S. Pearsons, and R. Bennet, "Prediction of Aural Detectability of Noise Signals," Human Factors, 16:4, 1974, 373-383.

Fletcher, H., "Auditory Patterns," Rev. Mod. Phys., 12, 1940, 47-65.

Freund, J. E., <u>Mathematical Statistics</u>, Englewood Cliffs, N. J.: Prentice-Hall, (1971).

Gilliom, J. D., C. Cline, and D. W. Taylor, "Harmonic Configuration: The Effect on the Hearing-Out Process," Paper delivered at the 93rd Meeting of the Acoustical Society of America, University Park, PA, June 1977. Abstract in J. Acoust. Soc. Amer., 61, 1977, S61.

Green, D. M., "Detection of Multiple Component Signals in Noise," J. Acoust. Soc. Amer., 30, 1958, 904-911.

a. Green, D. M., "Auditory Detection of a Noise Signal," J. Acoust. Soc. Amer., 32, 1960, 121-131.

b. Green, D. M., "Psychoacoustics and Detection Theory," J. Acoust. Soc. Amer., 32, 1960, 1189-1203.

Green, D. M., T. G. Birdsall, and W. P. Tanner, Jr., "Signal Detection as a Function of Signal Intensity and Duration," J. Acoust. Soc. Amer., 29:4, 1957, 523-531.

Green, D. M., and J. A. Swets, Signal Detection Theory and Psychophysics, New York: John Wiley and Sons, (1966).

Gundy, R. F., "Auditory Detection of an Unspecified Signal," J. Acoust. Soc. Amer., 33, 1961, 1008-1012.

Howard, J. H., "Psychological Structure of Eight Complex Underwater Sounds," J. Acoust. Soc. Amer., 62:1, 1977, 149-156.

Howard, J. H., and E. B. Silverman, "Structural Components in the Perception of Sixteen Complex Sounds," Human Performance Laboratory: Catholic University Technical Report ONR-75-1, August 1975.

Janota, C. P., "An Experimental Treatment of Auditory Discrimination of Complex Noise-Like Sounds," The Pennsylvania State University, Doctoral Dissertation, 1977.

Julesz, B., "Problems of Pattern Recognition and Perceptual Psychology," in L. N. Kanal (Ed.) Pattern Recognition, Washington, DC: Thompson Book Co., (1968).

Licklider, J. C. R., and D. M. Green, "Aural Detection of Complex Signals in Random Noise," in I. L. Cramer (Ed.) Proceedings of the Third International Congress on Acoustics, Amsterdam: Elsevier, (1961).

Massaro, D. W., "Preperceptual Images, Processing Time, and Perceptual Units in Auditory Perception," Psychol. Rev. 79:2, 1972, 124-145.

Mattingly, I. G., A. M. Liberman, A. K. Syrdal, and T. Halwes, "Discrimination in Speech and Nonspeech Modes," Cognitive Psychology, 2, 1971, 131-157.

Miller, G. A., "The Perception of Short Bursts of Noise," J. Acoust. Soc. Amer., 20:2, 1948, 160-170.

Miller, G. A., and W. G. Taylor, "The Perception of Repeated Bursts of Noise," J. Acoust. Soc. Amer., 20:2, 1948, 171-182.

Moore, B. C. J., and D. H. Raab, "Intensity Discrimination for Noise Bursts," J. Acoust. Soc. Amer., 57:2, 1975, 401-406.

Mundie, J. R., "Neurophysiological Principles of Auditory Information Processing," Aerospace Medical Research Laboratory Report AMRL-TR-68-186, 1970.

Nordmark, J. O., "Time and Frequency Analysis," in J. V. Tobias (Ed.) Foundations of Modern Auditory Theory, Vol. I, New York: Academic Press, (1972).

Norman, D. A., and D. G. Bobrow, "On Data-Limited and Resource-Limited Processes," Cognitive Psychology, 7, 1975, 44-64.

Oestreicher, H. L., "Natural and Artificial Auditory Information Processing," reprinted from Kybernetik, Muenchen-Wien: R. Oldenbourg Verlag, (1968).

Peters, R., "Perceptual Organization for Speech and Other Auditory Signals," Aerospace Medical Research Laboratory Report AMRL-TR-68-31, 1967.

Peterson, W. W., T. G. Birdsall, and W. C. Fox, "The Theory of Signal Detectability," Trans. I.R.E., PGIT-4, 1954, 171-212.

Plomp, R., "Pitch of Complex Tones," J. Acoust. Soc. Amer., 41:6, 1967, 1526-1533.

Raisbeck, G., Information Theory, An Introduction for Scientists and Engineers, Cambridge, Mass.: MIT Press, 1963.

Reed, S. K., Psychological Processes in Pattern Recognition, New York: Academic Press, (1973).

Ritsma, R. J., "Frequencies Dominant in the Perception of the Pitch of Complex Sounds," J. Acoust. Soc. Amer., 42:1, 1967, 191-198.

Robinson, D. E., and C. S. Watson, "Psychophysical Methods in Modern Psychoacoustics," in J. V. Tobias (Ed.) Foundations of Modern Auditory Theory, Vol. II, New York: Academic Press, (1972).

Scharf, B., "Critical Bands," in J. V. Tobias (Ed.) Foundations of Modern Auditory Theory, Vol. I, New York: Academic Press, (1972).

Shannon, C. E., and W. Weaver, The Mathematical Theory of Communication, Urbana, IL: University of Illinois Press, (1949).

Small, A. M., "Pure Tone Masking," J. Acoust. Soc. Amer., 31, 1959, 1619-1625.

Stallard, J. M., and C. E. Leslie, "Psychoacoustics and Passive Sonar Detection," Naval Ordnance Laboratory Technical Report, NOLTR 74-27, 1974.

Stevens, K. N., and A. S. House, "Speech Perception," in J. V. Tobias (Ed.) Foundations of Modern Auditory Theory, Vol. II, New York: Academic Press, (1972).

Stevens, S. S. (Ed.) Handbook of Experimental Psychology, New York: John Wiley and Sons, (1951).

Swets, J. A. (Ed.), Signal Detection and Recognition by Human Observers, New York: John Wiley and Sons, (1964).

Swets, J. A., and D. M. Green, "Sequential Observations by Human Observers of Signals in Noise," in C. Cherry (Ed.), Information Theory, London: Butterworths, (1961).

Swets, J. A., D. M. Green, and W. P. Tanner, Jr., "On the Width of Critical Bands," J. Acoust. Soc. Amer., 34, 1962, 108-113.

Swets, J. A., E. F. Shipley, M. J. Mikey, and D. M. Green, "Multiple Observations of Signals in Noise," J. Acoust. Soc. Amer., 31:4, 1959, 514-521.

Swets, J. A., W. P. Tanner, Jr., and T. G. Birdsall, "Decision Processes in Perception," Psychol. Rev., 68:5, 1961, 301-309.

Tanner, W. P., Jr., "Theory of Signal Detectability as an Interpretive Tool of Psychological Data," J. Acoust. Soc. Amer., 32:9, 1960, 1140-1147.

Tanner, W. P., Jr., and R. D. Sorkin, "The Theory of Signal Detectability," in J. V. Tobias (Ed.), Foundations of Modern Auditory Theory, Vol. II, New York: Academic Press, (1972).

Tartter, V. C., and P. D. Eimas, "The Role of Auditory Feature Detectors in the Perception of Speech," Perception and Psychophysics, 18:4, 1975, 293-298.

Warren, R. M., C. J. Obusek, R. M. Farmer, and R. P. Warren, "Auditory Sequence Confusion of Patterns Other than Speech or Music," Science, 164, 1969, 586-587.

Webster, J. C., M. M. Woodhead, and A. Carpenter, "Identifying Diesel Engine Sounds," J. Sound Vibration, 9:2, 1969, 241-246.

Webster, J. C., M. M. Woodhead, and A. Carpenter, "Perceptual Confusions Between Four-Dimensional Sounds," J. Acoust. Soc. Amer., 53, 1973, 448-456.

Woodhead, M. M., A. Carpenter, and J. C. Webster, "Learning to Identify Complex Sounds Using Explicit Instructions," Royal Navy Medical Research Council Report Br48923, January 1973.

Zagoruyko, N. G., and G. Ya. Voloshina (Ed.), "Auditory Pattern Recognition, USSR," translated from the Russian language text Raspoznavaniye Slukhovykh Obrazov, by the National Technical Information Service, Nauka Press, (1969).

Zhukov, S., and L. A. Christovitch, "Some New Concepts in Auditory Analysis of Acoustic Flow," Sov. Phys. Acoust., 20:3, 1974, 237-240.

APPENDIX A

INSTRUCTIONS TO SUBJECTS

Prior to the first experimental session, each subject was given a description of the objectives of the experiments, the experimental procedure to be used, and the rules regarding scheduling of sessions. Subjects were shown how to mount the tapes and operate the cassette machine used to record data for each session. They were encouraged to ask questions at any time during the seven months of the investigation. In addition, the following specific instructions appeared at the beginning of every audio tape to which the subjects listened.

The test sequence will consist of two signals presented without interfering noise. These signals will be denoted Signal A and Signal B. Signal A will be presented, then Signal B. The signals will then be repeated in the sequence A then B. During the response period, which will be indicated by a green light on the response recorder, either Signal A or Signal B will be presented in a noise. The amount of noise will decrease slowly. The objective is to indicate your decision as to which signal you conclude is mixed with the noise. Indicate your choice by pressing the switch marked A if you decide that Signal A was mixed with the noise, or, press the switch marked B if you decide that Signal B was mixed with the noise. You should

indicate this decision as soon as you can under the condition that you are reasonably certain of your choice. For this series of experiments, the tests are organized into groups of events. For all events of a group, the Signals A and B will be the same. The Signals A or B are presented randomly in the noise with each being equally likely. Now, please indicate your classification decision for the following cases. Voice comments following grouped events and leading into another group or terminating the test session.

This is the end of one group of tests. Another group of tests follows. For this group, the Signals A and B will be the same. However, these will generally not be the same signals as in the previous group. Please try to learn these signals without regard for other signals you have heard during this experiment.

This is the end of another group of tests. In the group of events which follows, the Signals A and B are the same throughout. These signals will generally not be the same as those heard previously. Please try to learn these sounds independent of other signals you may have been exposed to in this experiment.

End of another group of events. The final group of events follows. For this group as in those before, the Signals A and B will be the same. The likelihood of Signal A being presented in the noise is the same as is the likelihood of Signal B being presented in the noise.

This concludes a test session. Thank you for your cooperation.

APPENDIX B

ANALYSIS OF THE INTERDEPENDENCE OF SIGNAL-TO-NOISE RATIO AND TIME USING THE MODIFIED THRESHOLD PROCEDURE

The degree of correlation between signal-to-noise ratio and elapsed time from the beginning of a trial using the modified threshold procedure has been a matter of concern throughout this thesis. A high correlation between these variables prevents separate analysis of detectability-related and time-related factors which influence a subject's decision criterion. The modified threshold technique, being a sequential classification scheme, has the property that SNR is an increasing function of time. However, in an attempt to reduce the correlation between these variables, the starting values of SNR were randomized. As will be shown in the subsequent analysis, this randomization did not produce the desired effect.

In the experiments, the SNR started at some value S_0 and increased in 1/2-dB steps every two seconds. To facilitate further analysis, the quantization of this step function will be removed and replaced with a line of constant slope. That is,

$$\mathrm{SNR} = S_0 + (T/4). \tag{11}$$

This simplification will not greatly affect subsequent results since the continuous function used in the model exactly matches that in the actual experiment at the points of interest, the step changes.

Obviously, if the starting value S_0 were always constant, SNR would be completely dependent on time. The value of the correlation coefficient would be unity since SNR increases linearly with time. In order to reduce this dependence, the values of S_0 were randomized. The starting SNR for each event was chosen from the values -11, -10, -9, -8, and -7 dB, all of which occurred with an equal probability of 1/5. Thus, $f(S_0)$ is a discrete uniform distribution with

$$P(S_0) = 0.2 \text{ for } -11 \leq S_0 -7 \text{ dB, elsewhere.}$$
$$E(S_0) = 9 \text{ dB, and}$$
$$\mathrm{Var}(S_0) = 20/12 \quad \text{(Freund, 1971).} \tag{12}$$

In order to determine the correlation coefficient between SNR and time, the following parameters are needed: cov(SNR,T), var(SNR), and var(T) since the correlation coefficient is defined by:

$$\rho = \mathrm{cov}(SNR,T)/\left[\mathrm{Var}(SNR)\ \mathrm{var}(T))\right]^{1/2}. \tag{13}$$

Since it is not possible in this model to precisely define the distribution of time, an arbitrary distribution f(T) with variance σ_T^2 will be used. In order to determine the variance of SNR, it will be represented as the linear combination of two independent random variables, x and y, where $y = S_0$, and $x = T$. Then, $SNR = y+0.25*x$. The variables x and y are independent by design, i.e., S_0 is randomly chosen from its discrete uniform distribution, and T starts at zero and increases. The variance of a linear combination of two independent random variables can be determined as follows.

$$\mathrm{var}(a_1z_1+a_2z_2) = \sum_{i=1}^{2} a_i^2 \,\mathrm{var}(z_i) \tag{14}$$

where the a's are constants (Freund, 1971). Thus,

$$\mathrm{var}(y+.25x) = \mathrm{var}(y) + (\mathrm{var}(x)/16) = 20/12 + \sigma_T^2/16. \tag{15}$$

It will be of interest later to represent the variable y as having arbitrary limits instead of the -11 to -7 dB used in the experiments. Then,

$$\mathrm{var}(y) = k(k-1)/12$$

where k is the number of possible values which y can assume. For the experiments, k = 5. Using the above representation,

$$\mathrm{var}(SNR) = \left[k(k-1)/12\right] + \sigma_T^2/16. \tag{16}$$

To determine the covariance of SNR and T, the following theorem can be used:

If x1 and x2 are independent random variables, and a1, a2, b1, and b2 are constants such that SNR = a1x1+a2x2, and T = b1x1+b2x2, then

$$\mathrm{cov}(SNR,T) = \sum_{i=1}^{2} a_i b_i \,\mathrm{var}(x_i) \text{ (Freund, 1971).} \tag{17}$$

By definition, SNR = 1*y+0.25*x, and T = 0*y+1*x. The product a1*b1 is therefore zero, and a2*b2 = 0.25. The covariance of SNR and T is then 0.25*var(x), and since x and T are just different names for the same variable,

$$\text{cov}(\text{SNR},T) = 0.25\ \sigma_T^2. \tag{18}$$

The correlation coefficient is then

$$\rho = 0.25\ \sigma_T^2 / \sqrt{\left[\sigma_T^2\ k(k-1)/12\right] + \sigma_T^2/16}$$

$$= \sigma_T / \sqrt{\sigma_T^2 + 4k(k-1)/3}. \tag{19}$$

For the present experiments with k = 5,

$$\rho = \sigma_T / \sqrt{\sigma_T^2 + 26.67}\ . \tag{20}$$

From the experimental data, the sample standard deviation of the distribution of response time is on the order of 10 to 20 seconds. With k = 5 and σ_T = 10 seconds, the value of the correlation coefficient is 0.845. Larger values of σ_T produce even higher values of the correlation coefficient. Thus, it can be clearly seen that SNR and T are very much interdependent. In order to decrease this correlation, either the variance in response time must be greatly reduced, or the range of starting SNR values must be greatly increased. The variance in response time, however, is a function of the listener's consistency and is therefore not a possible control within the experimental design. Using Equation (19), it can be shown that in order to reduce the correlation coefficient below 0.5, the starting SNR values would have to range over some 15 dB. This is, of course, not practical in the modified threshold procedure. It can be shown that the correlation is reduced slightly by reducing the rate of

increase of SNR with time below 0.25 dB/sec. However, this does not produce satisfactory results since by this method, trials would become so long that subjects would not remember the signal characteristics.

Finally, it must be pointed out that a low correlation does not necessarily indicate independence of the parameters, unless the distributions are assumed to be Gaussian. Clearly, and most unfortunately, randomization of the starting values of SNR has not greatly reduced the dependence of SNR on time. Consequently, criterion effects based on these variables cannot be quantitatively analyzed as separate factors using the modified threshold procedure.

DISTRIBUTION

Commander (NSEA 09G32) Naval Sea Systems Command Department of the Navy Washington, D. C. 20362	Copies 1 and 2
Commander (NSEA 0342) Naval Sea Systems Command Department of the Navy Washington, D. C. 20362	Copies 3 and 4
Defense Documentation Center 5010 Duke Street Cameron Station Alexandria, VA 22314	Copies 5 through 16